

2007/80



Testing fiscal sustainability in Poland:
a Bayesian analysis of cointegration

Andrea Silvestrini

CORE DISCUSSION PAPER
2007/80

**Testing fiscal sustainability in Poland:
a Bayesian analysis of cointegration**

Andrea SILVESTRINI¹

October 2007

Abstract

Fiscal sustainability is a central topic for most of the transition economies of Eastern Europe. This paper focuses on a particular country: Poland. The main purpose is to investigate, empirically, whether the post-transition fiscal policy is consistent with the intertemporal budget constraint, used as a formal theoretical framework. To test debt stabilization, the empirical analysis is made in two steps in which different inferential approaches are adopted. In the *first* step we perform the preliminary unit roots analysis and the selection of the cointegration rank using parametric and bootstrap procedures. In the *second* step we apply Bayesian inference to the estimation of the cointegrating vector and of the adjustment parameters. In this way, we experiment the usefulness of Bayesian inference in precisely assessing the magnitude of the cointegrating vector. Moreover, we show to what extent the likelihood of the data is important in revising the available prior information, relying on numerical integration techniques.

Keywords: Bayesian inference, fiscal sustainability, cointegration, bootstrap.

JEL Classification: C11, C32, E62

¹ CORE, Université catholique de Louvain, Belgium and Bank of Italy, Research Department, Roma, Italy.

The author is indebted to Luc Bauwens for helpful comments on an earlier draft. The author wishes to thank Pierluigi Daddi, Joanna Mlynarczyk, Gianluca Moretti, Giulio Nicoletti and David Veredas for useful discussion and suggestion. The views expressed in the paper are those of the author and do not necessarily represent those of the Bank of Italy.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the author.

1 Introduction

For many industrialized countries, concerns over the feasibility of a permanent deficit lay stress on the issue of fiscal sustainability, meaning whether a government can continue to operate under its current fiscal policy indefinitely without any change. In the European Union, the adoption of the Stability and Growth Pact has raised the importance of monitoring budgetary developments in the member countries. Fiscal sustainability criteria have been introduced as rule-based control for the eligibility of non-member countries in Central and Eastern Europe. Consequently, this issue has received increasing attention from economists and policy makers.

From the macroeconomic literature¹, it is well known that for the government to be solvent the real debt has to grow less rapidly than the real interest rate: This is equivalent to fiscal sustainability. Empirically, to explore whether the solvency condition is fulfilled, several studies have performed stationarity tests on the discounted real debt series. For how long can government budget deficits continue unchecked? What are the limits on a government's actions? Hamilton and Flavin (1986) pose these questions in their seminal paper. To answer them, they derive the government's intertemporal budget constraint. Compliance with this constraint implies that the government is setting the current market value of the debt equal to the discounted sum of expected future surpluses. According to these authors, assuming a constant real interest rate, this can be checked by running a battery of stationarity tests on the real debt and on the real primary surplus.

Trehan and Walsh (1988) point out that, under the assumption of a constant real interest rate, the stationarity test on the discounted debt is equivalent to the cointegration test between government expenditures inclusive of interest, tax receipts and seignorage. Wilcox (1989) extends Hamilton and Flavin findings by allowing for stochastic real interest rates and proposing an alternative test. Stationarity of the discounted debt is studied by estimating the coefficients of a univariate autoregressive representation of the series: This is a way to deal with the problem of residuals serial correlation when carrying out unit root tests. Trehan and Walsh (1991) examine what happens if the expected real rate of interest is no longer constant and inquire the validity of the cointegration test in distinguishing sustainable policies from unsustainable ones. The same authors highlight (p.207) that the majority of the papers focusing on the empirical consistency of intertemporal budget balance "Develop their tests by exploiting the presence, under intertemporal budget balance, of a cointegrating relationship linking net-of-interest expenditures, revenues, interest payments and the outstanding stock of the debt". Cointegration is therefore a useful tool for the study of fiscal sustainability.

Hakkio and Rush (1991) test for cointegration between revenues (*rev*) and expenditures inclusive of interest payments (*exp*) under the assumption that the stochastic real interest rate is stationary and that revenues and expenditures are non-stationary. Remarkably, they state two necessary conditions for compliance with the intertemporal budget constraint: a) cointegration of revenues and expenditures with b) cointegrating vector equal to $[-1 \ \beta_{exp}]'$ and $0 < \beta_{exp} \leq 1$, normalizing on revenues. However, they show that, if $\beta_{exp} < 1$, the limit of the undiscounted value of the debt is infinity and positive shocks to expenditures generate permanently lower revenues through the cointegrating relationship. As a consequence, governments experience growing unease in marketing their mounting debts and presumably face a greater incentive to default. Therefore, the cointegrating vector must be equal to $[-1 \ 1]$ to rule out this possibility.

Haug (1991) tests for unit roots and cointegration of real debt and real primary surplus variables, calculating the finite sample critical values via Monte Carlo methods. Quintos (1995) introduces 'strong' and 'weak' conditions for fiscal sustainability to hold. The strong one is equivalent to cointegration between revenues and expenditures as a necessary and sufficient condition for sustainability. The weak one requires (as necessary and sufficient condition) the debt process to grow slower than the growth rate of the mean interest rate. These conditions are also discussed by Martin (2000).

¹See, for instance, Ljungqvist and Sargent (2004) and Buiter (2005).

Despite the large number of empirical and theoretical papers focusing on fiscal sustainability in the US and in Western European countries², similar investigations on developing and transition economies have been by far less numerous. Among these, Kalyoncu (2005) examines sustainability of fiscal stances in five developing countries (Mexico, South Africa, the Philippines, South Korea and Turkey). Buiter (1996) and Budina and van Wijnbergen (1997) constitute representative examples of studies concerning fiscal solvency in Eastern Europe. More recently, Green *et al.* (2001) evaluate deficit sustainability in Poland between 1991 and 1998.

Our analysis, similarly to Green *et al.* (2001), focuses on Poland. This choice is not random. Our motivation partly stems from the fact that this was one of the first countries to start the transition process to a market economy, in 1989. As such, it has one of the longest histories of transition experience³. Undoubtedly, therefore, Poland offers a relatively favorable institutional setting within which to study fiscal sustainability.

In this paper, we examine whether current fiscal policy in Poland satisfies the intertemporal budget constraint. We start from the government's one-period budget constraint, which is equation (1) in Arghyrou and Luintel (2007). By taking expectations with respect to the information set at t and by solving equation (1) forwardly, Arghyrou and Luintel get to equation (2), that we list here below for the sake of clarity

$$GB_t = \mathbb{E}_t \left[\sum_{j=0}^{\infty} (1+R)^{-(j+1)} (rev_{t+j} - exp_{t+j}) \right] + \lim_{j \rightarrow \infty} \mathbb{E}_t \left[(1+R)^{-(j+1)} GB_{t+j+1} \right],$$

where GB_t is the stock of government real debt at time t , rev_t and exp_t are real revenues and real expenditures at time t , R is the real interest rate (whose conditional expectation is assumed to be constant) and \mathbb{E}_t is the expected value operator conditional on the information set available at t . To rule out *Ponzi games*, i.e. to avoid "The possibility of the government financing old debt that matures by issuing new debt" (Hakkio and Rush, 1991, p.431), it must hold that the limit of the expected value of the discounted future stock of real debt goes to zero:

$$\lim_{j \rightarrow \infty} \mathbb{E}_t \left[(1+R)^{-(j+1)} GB_{t+j+1} \right] = 0. \quad (1)$$

In this way the intertemporal budget constraint is satisfied and the sustainability hypothesis holds. In our work we investigate whether the Polish data are consistent⁴ with (1) over the sample period. To this aim, a methodology suggested by the influential literature is to test for cointegration between revenues and expenditures and to estimate the parameter β_{exp} in the normalized cointegrating vector $[-1 \ \beta_{exp}]'$. We refer to Westerlund and Prohl (2007), among others, for further insights.

²For the US, we refer to Trehan and Walsh (1988), Wilcox (1989), Haug (1991) and Hakkio and Rush (1991). For Western European countries, see among others Corsetti and Roubini (1991) for some tests of government solvency applied to a sample of 18 OECD countries, Vanhorebeek and Van Rompuy (1995) for an analysis of sustainability over a sample of eight European countries, Caporale (1995) for a similar study conducted on ten EU member countries and Payne (1997) for an analysis of sustainability of G-7 countries. Moreover, Apergis *et al.* (2000) focus on the current account deficit in Greece, Greiner *et al.* (2006) apply a test for sustainability to German data, Haber and Neck (2006) investigate sustainability of Austrian public debt while Fonseca Marinheiro (2006) examines Portuguese fiscal policies over a long time horizon. Finally, we refer to Arghyrou and Luintel (2007) for a very recent re-examination of Corsetti and Roubini's data set and to Westerlund and Prohl (2007) for new evidence (based on panel data) of fiscal sustainability within the European Union.

³The transition path to capitalism has revealed itself to be much more complex than expected. Nevertheless, once this process was undertaken, Poland moved faster than other countries by implementing a number of fiscal reforms aiming to reduce the tax burden, improving the allocation of budget resources and approaching the EU standards.

⁴We are implicitly assuming that the fiscal policy remains unchanged. In other terms, as pointed out by Hakkio and Rush (1991), the hypothesis is that total revenues and expenditures "Continue to follow the stochastic process that we estimate" (p.431).

In a nutshell, therefore, what we do is to test 1) cointegration between revenues and expenditures (inclusive of interest) and to estimate 2) the posterior distribution of parameter β_{exp} in the normalized cointegrating vector $[-1 \ \beta_{exp}]'$. To begin with, the preliminary unit roots analysis and the selection of the cointegration rank are carried out within a non-Bayesian inferential setting. In a second step, Bayesian inference is applied to derive the posterior distribution of the cointegrating vector and to calculate its first two moments. With a single cointegrating relation, we make use of some known results concerning the posterior density of the cointegrating vector, which belongs to the poly-t densities class. We also show to what extent the likelihood of the data is important in revising the available prior information. To do this, we rely on numerical integration techniques based on deterministic methods.

The main contribution of the paper is the empirical problem to be analyzed and the kind of econometric strategy employed. On a theoretical ground, the methodologies adopted are all taken from the recent literature. Dummy variables are introduced in the deterministic part of the model to deal with the seasonal nature of the data. The unit root testing is carried out both in a parametric both in a non-parametric framework: The results derived by using the parametric approach are carefully compared to those of alternative bootstrap procedures. In this way, we believe that a more appropriate test decision may be taken. The Bayesian analysis of cointegration, with a transparent choice of the prior and with easy-to-implement posterior calculations, enables us to conclude tentatively whether and to what extent fiscal sustainability can be achieved in the long-run.

Notice that all the papers so far cited comprise non-Bayesian analyses of cointegration. However, as pointed out by Villani (2005), p.327, "Philosophical issues aside, a Bayesian approach is advantageous" since "It produces whole probability distributions for each unknown parameter that are valid for any sample size". This is the gain of the Bayesian approach with respect to the frequentist. To our knowledge, fiscal sustainability is an issue that has never been tackled within the Bayesian framework. This article is an attempt to fill this gap.

The paper is organized as follows. In Section 2 we introduce some notation and quickly review cointegrated VAR systems. In Section 3 we outline a Bayesian analysis of cointegrated VAR models. Section 4 contains a preliminary empirical analysis concerning data issues, unrestricted VAR estimation, unit root tests and selection of the cointegration rank. In Section 5 we apply Bayesian inference to assess the sustainability of fiscal policy in Poland. Section 6, finally, draws our conclusions. The proofs are gathered in the Appendix.

2 Setup: Cointegrated VAR models

This Section briefly reviews unrestricted VAR models and cointegrating VAR systems. To start, let us introduce the general VAR of order p as

$$\mathbf{y}_t = \mathbf{\Theta} \mathbf{D}_t + \sum_{i=1}^p \mathbf{A}_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (2)$$

where $\{\mathbf{y}_t\}_{t=1}^T$ is a n -dimensional vector of variables, \mathbf{A}_i ($i = 1, \dots, p$) are $n \times n$ matrices and p is the maximum lag length, assumed to have already been selected. Innovations are $\boldsymbol{\varepsilon}_t \sim N_n(0, \boldsymbol{\Sigma})$, with positive definite covariance matrix $\boldsymbol{\Sigma}$ and independent over time. Moreover, $\mathbf{\Theta}$ is a $n \times m$ matrix for the m deterministic variables in \mathbf{D}_t . The $m \times 1$ vector of coefficients \mathbf{D}_t may contain, for instance, a constant, a linear term and seasonal dummies.

The VAR in equation (2) can be parameterized in vector error correction (VECM) form as

$$\Delta \mathbf{y}_t = \mathbf{\Theta} \mathbf{D}_t + \sum_{i=1}^{p-1} \boldsymbol{\Pi}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T-1, \quad (3)$$

where $\mathbf{\Pi} = -\mathbf{A}(1) = -\left(\mathbf{I}_n - \sum_{i=1}^p \mathbf{A}_i\right)$ is the matrix which describes the long-run relationships among the variables, while $\mathbf{\Pi}_i = \sum_{j=i+1}^p \mathbf{A}_j$, $(i = 1, \dots, p-1)$ are the matrices accounting for the effects of short-run dynamics.

When the vector \mathbf{y}_t of n random variables in (2) is integrated of order one ($I(1)$) and yet for sure one or several linear combinations of these variables are integrated of order zero ($I(0)$), we are dealing with a cointegrated VAR system. These variables are said to be cointegrated. According to the Granger's Representation Theorem, under a set of assumptions, the n -dimensional vector \mathbf{y}_t may be represented as the sum of a random walk and a stationary process (see Johansen, 1996).

In a cointegrated VAR system, the rank of $\mathbf{\Pi}$ matrix in (3) is equal to the number of independent cointegrating vectors. If the rank of $\mathbf{\Pi}$ is r and $r < n$, $\mathbf{\Pi}$ has reduced rank and there are $r < n$ linearly independent combinations of \mathbf{y}_t that are stationary. Therefore, the variables in \mathbf{y}_t are $I(1)$ with r cointegrating vectors and $n - r$ unit roots.

When $\mathbf{\Pi}$ has reduced rank, it is possible to decompose it in $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices of rank r . This rank restriction may be incorporated in the VECM re-writing (3) as⁵

$$\Delta \mathbf{y}_t = \boldsymbol{\Theta} \mathbf{D}_t + \sum_{i=1}^{p-1} \mathbf{\Pi}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (4)$$

where $\boldsymbol{\beta}' \mathbf{y}_{t-1}$ constitutes the error correction term and the equation $\boldsymbol{\beta}' \mathbf{y}_t = 0$ represents the long-run equilibrium model for the components of \mathbf{y}_t . Inference on $\boldsymbol{\beta}$ matrix is an issue of great empirical relevance: $\boldsymbol{\beta}' \mathbf{y}_t = 0$ equation, indeed, may be viewed as a long-run equilibrium model linking the components of \mathbf{y}_t . Moreover, the elements of $\boldsymbol{\alpha}$ are known as the adjustment parameters in the VECM model, measuring how quickly deviations from long-run equilibrium feed back into the system. The larger their absolute value, the quicker the adjustment of the system towards its long-run equilibrium path.

The parametrization $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ implies identification issues on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. These matrices are not uniquely identified since $\mathbf{\Pi}$ can be factorized as $(\boldsymbol{\alpha}\mathbf{H}^{-1})(\mathbf{H}\boldsymbol{\beta}') = \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\beta}'}$, choosing any full rank matrix \mathbf{H} (of dimension $r \times r$). According to Bauwens and Lubrano (1996), r^2 restrictions for identification have to be imposed on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. A practical solution in terms of economic interpretation is to set restrictions on $\boldsymbol{\beta}$, leaving $\boldsymbol{\alpha}$ unrestricted. Linear restrictions on $\boldsymbol{\beta}$ can be introduced defining

$$\underset{(n \times r)}{\boldsymbol{\beta}} = \begin{bmatrix} -\mathbf{I}_r \\ \boldsymbol{\beta}^* \end{bmatrix}, \quad \underset{((n-r) \times r)}{\boldsymbol{\beta}^*} \quad (5)$$

where $\boldsymbol{\beta}^*$ is the unrestricted matrix. In what follows we shall stick to this definition and we shall keep $\boldsymbol{\alpha}$ unrestricted.

3 Bayesian inference: Model, priors and posteriors

Bayesian inference on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ is hereafter reviewed following the analysis conducted by Bauwens *et al.* (1999). Regarding the likelihood function, the prior and the posterior density, results are summarized for the general case and for the case of a single cointegrating vector. Note that this is *one* of the proposals

⁵Notice that we do not distinguish between deterministic variables that are restricted to lie in the cointegration space and those that are not: From here onwards, we assume that the deterministic variables all lie outside the cointegration space.

presented in the Bayesian literature, see for instance Kleibergen and van Dijk (1994), Geweke (1996), Kleibergen and Paap (2002), Strachan (2003), Villani (2005) and the references therein. See Warne (2006) for a very recent application of Bayesian techniques in cointegrated VAR models to the estimation of euro area demand for M3.

The starting point is expressing the VECM in (4) in matrix form as a regression model

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}\mathbf{\Gamma} + \mathbf{Z}\mathbf{\beta}\mathbf{\alpha}' + \mathbf{E} \\ &= \begin{bmatrix} \mathbf{X} & \mathbf{Z}\mathbf{\beta} \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{\alpha}' \end{bmatrix} + \mathbf{E} \\ &= \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{Z}} \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{\alpha}' \end{bmatrix} + \mathbf{E},\end{aligned}\tag{6}$$

where

$$\underset{(T \times n)}{\mathbf{Y}} = \begin{bmatrix} \Delta \mathbf{y}'_1 \\ \vdots \\ \Delta \mathbf{y}'_T \end{bmatrix}, \quad \underset{(T \times n)}{\mathbf{E}} = \begin{bmatrix} \boldsymbol{\varepsilon}'_1 \\ \vdots \\ \boldsymbol{\varepsilon}'_T \end{bmatrix}, \quad \underset{(T \times n)}{\mathbf{Z}} = \begin{bmatrix} \mathbf{y}'_0 \\ \vdots \\ \mathbf{y}'_{T-1} \end{bmatrix}.$$

Moreover, \mathbf{X} and $\mathbf{\Gamma}$ correspond to

$$\underset{(T \times (m+n(p-1)))}{\mathbf{X}} = \begin{bmatrix} \mathbf{D}'_0 & \Delta \mathbf{y}'_0 & \cdots & \Delta \mathbf{y}'_{2-p} \\ \mathbf{D}'_1 & \Delta \mathbf{y}'_1 & \cdots & \Delta \mathbf{y}'_{3-p} \\ \vdots & \vdots & & \vdots \\ \mathbf{D}'_{T-1} & \Delta \mathbf{y}'_{T-1} & \cdots & \Delta \mathbf{y}'_{T-p+1} \end{bmatrix}, \quad \underset{((m+n(p-1)) \times n)}{\mathbf{\Gamma}} \begin{bmatrix} \boldsymbol{\Theta}' \\ \boldsymbol{\Pi}'_1 \\ \vdots \\ \boldsymbol{\Pi}'_{p-1} \end{bmatrix}.\tag{7}$$

Note that $\mathbf{\beta}$ and $\mathbf{\alpha}$ comprise the parameters of interest. The matrices $\boldsymbol{\Theta}'$, $\boldsymbol{\Pi}'_1$, \dots , $\boldsymbol{\Pi}'_{p-1}$ and $\boldsymbol{\Sigma}$ contain nuisance parameters.

Result 3.1 *The model in (6) may be expressed in compact form as*

$$\mathbf{Y} = \tilde{\mathbf{W}}\mathbf{B} + \mathbf{E},\tag{8}$$

where $\tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{X} & \mathbf{Z}\mathbf{\beta} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \mathbf{\Gamma} & \mathbf{\alpha}' \end{bmatrix}'$. Once we condition on $\mathbf{\beta}$, the model in (8) is linear in its parameters. Hence, it is possible to make use of all the known results for multivariate regression models in the Bayesian framework.

See Appendix A.1 for the algebraic details.

The likelihood function⁶ that corresponds to the stacked regression in (8) is

$$\mathcal{L}(\mathbf{\beta}, \mathbf{B}, \boldsymbol{\Sigma} \mid \text{Data}) \propto |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr } \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B})' (\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B}) \right\},\tag{9}$$

where $\text{Data} = \{\mathbf{Y}, \mathbf{X}, \mathbf{Z}\}$ indicates the available sample information.

As pointed out by Warne (2006), the choice of a proper prior distribution of the parameters is a step of paramount importance. Several alternatives are at hand. See, among others, Koop *et al.* (2006) for

⁶The symbol $\mathcal{L}(\cdot)$ denotes the likelihood function of generic argument (\cdot) .

an updated survey. Following Villani (2005), we can decompose the general prior distribution on the parameters as

$$\phi(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma}, r) = \phi(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma} \mid r) \times \phi(r),$$

where $\phi(r)$ is a prior density for the cointegration rank⁷.

In what follows we assume the cointegrating rank to be already known or estimated. Moreover, as a prior information on \mathbf{B} and $\boldsymbol{\Sigma}$, we introduce the standard diffuse prior

$$\phi(\mathbf{B}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{(n+1)}{2}}. \quad (10)$$

Strachan and Inder (2004) use a different approach, defining a weakly informative prior density on $\boldsymbol{\alpha}$. A different route is also taken by Strachan and van Dijk (2003), Villani (2005) and Warne (2006). We prefer to remain non-informative on these parameters since we have no prior information on them.

On the contrary, we set an informative prior on the unrestricted elements of $\boldsymbol{\beta}$ vector. Any prior density on the unrestricted elements of $\boldsymbol{\beta}$ can be used. We denote it with $\phi(\boldsymbol{\beta})$. We use $\boldsymbol{\beta}$ as an argument of the density function, even if some elements are known with probability one (whenever we impose some normalization restrictions, as in (5)). As a result, the joint prior information is given by the product

$$\phi(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma}) \propto \phi(\boldsymbol{\beta}) \times |\boldsymbol{\Sigma}|^{-\frac{(n+1)}{2}}. \quad (11)$$

The joint posterior density of $(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma})$ is equal to the product of the likelihood function $\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma} \mid \text{Data})$ in (9) times the joint prior $\phi(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma})$ in (11), i.e.

$$\phi(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma} \mid \text{Data}) \propto \phi(\boldsymbol{\beta}) \times |\boldsymbol{\Sigma}|^{-\frac{(T+n+1)}{2}} \exp \left\{ -\frac{1}{2} \text{tr } \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B})' (\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B}) \right\}. \quad (12)$$

Let us define $\mathbf{E} = (\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B})' (\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B})$. Thus, the joint posterior density $\phi(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma} \mid \text{Data})$ in (12) may be expressed as

$$\phi(\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\Sigma} \mid \text{Data}) \propto \phi(\boldsymbol{\beta}) \times |\boldsymbol{\Sigma}|^{-\frac{(T+n+1)}{2}} \exp \left\{ -\frac{1}{2} \text{tr } \boldsymbol{\Sigma}^{-1} (\mathbf{E}'\mathbf{E}) \right\}. \quad (13)$$

Note that, given $\boldsymbol{\beta}$ and \mathbf{B} , equation (13) is proportional to the kernel of an inverted Wishart distribution with parameters $\mathbf{E}'\mathbf{E}$ (positive definite) and $T > n - 1$ degrees of freedom⁸, i.e.

$$\phi(\boldsymbol{\Sigma} \mid \boldsymbol{\beta}, \mathbf{B}, \text{Data}) \propto |\boldsymbol{\Sigma}|^{-\frac{(T+n+1)}{2}} \exp \left\{ -\frac{1}{2} \text{tr } (\boldsymbol{\Sigma}^{-1} \mathbf{E}'\mathbf{E}) \right\}, \quad (14)$$

therefore $\boldsymbol{\Sigma} \mid (\boldsymbol{\beta}, \mathbf{B}, \text{Data}) \sim IW_n(\mathbf{E}'\mathbf{E}, T)$. Integrating (14) with respect to $\boldsymbol{\Sigma}$, we get the inverse of the integrating constant of the inverted Wishart distribution,

$$\begin{aligned} \int \phi(\boldsymbol{\Sigma} \mid \boldsymbol{\beta}, \mathbf{B}, \text{Data}) d\boldsymbol{\Sigma} &= \int |\boldsymbol{\Sigma}|^{-\frac{(T+n+1)}{2}} \exp \left\{ -\frac{1}{2} \text{tr } (\boldsymbol{\Sigma}^{-1} \mathbf{E}'\mathbf{E}) \right\} d\boldsymbol{\Sigma} \\ &= C(n, T) \times |\mathbf{E}'\mathbf{E}|^{-\frac{1}{2}T}, \end{aligned}$$

where

$$C(n, T) = 2^{\frac{1}{2}Tn} \pi^{\frac{1}{4}n(n-1)} \prod_{i=1}^n \Gamma \left(\frac{T+1-i}{2} \right)$$

⁷For instance, Strachan and Inder (2004) define for the cointegration rank $\phi(r) = (n+1)^{-1}$ as a prior density.

⁸We refer to Box and Tiao (1973) for a definition and properties of this distribution.

is a constant that does not depend on β and \mathbf{B} ⁹. Therefore the joint posterior density of β and \mathbf{B} is

$$\phi(\beta, \mathbf{B} \mid \text{Data}) \propto \phi(\beta) \times |\mathbf{E}' \mathbf{E}|^{-\frac{1}{2}T}. \quad (15)$$

Result 3.2 *We can express (15) as*

$$\phi(\beta, \mathbf{B} \mid \text{Data}) \propto \phi(\beta) \times |(\tilde{\mathbf{S}} + (\mathbf{B} - \tilde{\mathbf{B}})' \tilde{\mathbf{W}}' \tilde{\mathbf{W}} (\mathbf{B} - \tilde{\mathbf{B}}))|^{-\frac{1}{2}T}, \quad (16)$$

where $\tilde{\mathbf{B}} = (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}$ and $\tilde{\mathbf{S}} = (\mathbf{Y} - \tilde{\mathbf{W}} \tilde{\mathbf{B}})' (\mathbf{Y} - \tilde{\mathbf{W}} \tilde{\mathbf{B}})$.

For a proof see Appendix A.2.

In (16), given β , we can recognize the kernel of a matrixvariate Student distribution for \mathbf{B} (see again Appendix A in Bauwens *et al.*, 1999), therefore

$$\phi(\mathbf{B} \mid \beta, \text{Data}) \propto |(\tilde{\mathbf{S}} + (\mathbf{B} - \tilde{\mathbf{B}})' \tilde{\mathbf{W}}' \tilde{\mathbf{W}} (\mathbf{B} - \tilde{\mathbf{B}}))|^{-\frac{1}{2}T}, \quad (17)$$

and $\mathbf{B} \mid \beta, \text{Data} \sim Mt_{(m+n(p-1)+r) \times n}(\tilde{\mathbf{B}}, \tilde{\mathbf{W}}' \tilde{\mathbf{W}}, \tilde{\mathbf{S}}, T - (m + n(p-1)) - r)$.

The marginal posterior density of β is obtained by integrating with respect to \mathbf{B} equation (16)

$$\begin{aligned} \phi(\beta \mid \text{Data}) &= \int \phi(\beta, \mathbf{B} \mid \text{Data}) d\mathbf{B} \propto \phi(\beta) \times \int |(\tilde{\mathbf{S}} + (\mathbf{B} - \tilde{\mathbf{B}})' \tilde{\mathbf{W}}' \tilde{\mathbf{W}} (\mathbf{B} - \tilde{\mathbf{B}}))|^{-\frac{1}{2}T} d\mathbf{B} \\ &= \phi(\beta) \times C_{Mt}(\tilde{\mathbf{B}}, \tilde{\mathbf{W}}' \tilde{\mathbf{W}}, \tilde{\mathbf{S}}) \times |\tilde{\mathbf{S}}|^{-\frac{(T-k-r)}{2}} |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-\frac{n}{2}}, \end{aligned} \quad (18)$$

where for the ease of the exposition we let $k = m + n(p-1)$.

In (18), $C_{Mt}(\tilde{\mathbf{B}}, \tilde{\mathbf{W}}' \tilde{\mathbf{W}}, \tilde{\mathbf{S}})$ is the inverse of the integrating constant of the matrixvariate Student distribution and is independent on β ,

$$C_{Mt}(\tilde{\mathbf{B}}, \tilde{\mathbf{W}}' \tilde{\mathbf{W}}, \tilde{\mathbf{S}}) = \pi^{\frac{1}{2}n(k+r)} \prod_{i=1}^n \frac{\Gamma\left(\frac{T-k-r+1-i}{2}\right)}{\Gamma\left(\frac{T+1-i}{2}\right)}.$$

Therefore

$$\phi(\beta \mid \text{Data}) \propto \phi(\beta) \times |\tilde{\mathbf{S}}|^{-\frac{(T-k-r)}{2}} |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-\frac{n}{2}}. \quad (19)$$

Result 3.3 *Defining a non-informative prior on \mathbf{B} and Σ as in (10), an informative prior on β and keeping α unrestricted, it is possible to show that the marginal posterior density of β in (19) can be expressed as*

$$\phi(\beta \mid \text{Data}) \propto \phi(\beta) \times \frac{|\beta' \mathbf{W}_2 \beta|^{l_2}}{|\beta' \mathbf{W}_1 \beta|^{l_1}}, \quad (20)$$

where

$$\begin{aligned} \mathbf{W}_2 &= \mathbf{Z}' \mathbf{M}_X \mathbf{Z} \\ \mathbf{W}_1 &= \mathbf{Z}' \mathbf{M}_Y \left(\mathbf{I}_T - \mathbf{X}(\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{M}_Y \mathbf{Z}, \end{aligned} \quad (21)$$

$l_2 = \frac{T-k-r-n}{2}$, $l_1 = \frac{T-k-r}{2}$ and $\mathbf{M}_Y = \mathbf{I}_T - \mathbf{Y}(\mathbf{Y}' \mathbf{Y})^{-1} \mathbf{Y}'$.

⁹Moreover, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx$, $x > 0$, by definition of the Gamma function.

For a complete proof¹⁰ of this result see Appendix A.3.

We remind that r is the number of linear restrictions and $k = m + n(p - 1)$ is the number of columns of the matrix \mathbf{X} . The expression on the right-hand side of (20) is the product of a Student kernel times a ratio of determinants involving two quadratic forms in $\boldsymbol{\beta}$. The resulting density $\phi(\boldsymbol{\beta} | \text{Data})$ does not belong to a known class of distributions. Ideally, we should be able to recover the marginal density of any element of $\boldsymbol{\beta}$ and to calculate the moments, integrating (20). Unfortunately, dealing with several cointegrating vectors, no analytical results can be obtained. Some analytical results, however, are available in the case of a single cointegrating vector. We refer, again, to Bauwens *et al.* (1999).

Finally, we derive results for the $\mathbf{B}_{((k+r) \times n)} = \begin{bmatrix} \boldsymbol{\Gamma} \\ \boldsymbol{\alpha}' \end{bmatrix}$ matrix. From (17), we know that

$$\mathbf{B} | \boldsymbol{\beta}, \text{Data} \sim Mt_{(k+r) \times n}(\tilde{\mathbf{B}}, \tilde{\mathbf{W}}' \tilde{\mathbf{W}}, \tilde{\mathbf{S}}, T - k - r),$$

where $\tilde{\mathbf{B}} = (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}$, $\tilde{\mathbf{S}} = (\mathbf{Y} - \tilde{\mathbf{W}} \tilde{\mathbf{B}})' (\mathbf{Y} - \tilde{\mathbf{W}} \tilde{\mathbf{B}})$ and $\tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \boldsymbol{\beta} \end{bmatrix}$.

Hence $\boldsymbol{\alpha}'$ also follows a matrixvariate Student distribution. Consequently, conditionally on $\boldsymbol{\beta}$, each element of $\boldsymbol{\alpha}'$ has a univariate Student distribution (this can be proved applying results on matrixvariate Student distributions). The first moment of \mathbf{B} is

$$\begin{aligned} \mathbb{E}(\mathbf{B} | \boldsymbol{\beta}, \text{Data}) &= \tilde{\mathbf{B}} \\ &= \begin{bmatrix} \mathbf{X}' \mathbf{X} & \mathbf{X}' \mathbf{Z} \boldsymbol{\beta} \\ \boldsymbol{\beta}' \mathbf{Z}' \mathbf{X} & \boldsymbol{\beta}' \mathbf{Z}' \mathbf{Z} \boldsymbol{\beta} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \\ \boldsymbol{\beta}' \mathbf{Z}' \end{bmatrix} \mathbf{Y} \\ &= \begin{bmatrix} (\mathbf{X}' \mathbf{X})^{-1} + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \boldsymbol{\beta} \mathbf{F}_2 \boldsymbol{\beta}' \mathbf{Z}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} & -(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \boldsymbol{\beta} \mathbf{F}_2 \\ -\mathbf{F}_2 \boldsymbol{\beta}' \mathbf{Z}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} & \mathbf{F}_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}' \\ \boldsymbol{\beta}' \mathbf{Z}' \end{bmatrix} \mathbf{Y}, \end{aligned} \quad (22)$$

for $T - k - r > n$ and where $\mathbf{F}_2 = (\boldsymbol{\beta}' \mathbf{Z}' \mathbf{Z} \boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{Z}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \boldsymbol{\beta})^{-1}$. Let us focus on the partition corresponding to $\boldsymbol{\alpha}'$. It holds that

$$\begin{aligned} \tilde{\mathbf{B}} &= \begin{bmatrix} -\mathbf{F}_2 \tilde{\mathbf{Z}}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \mathbf{F}_2 \tilde{\mathbf{Z}}' \end{bmatrix} \mathbf{Y} \\ &= \begin{bmatrix} \mathbf{F}_2 \tilde{\mathbf{Z}}' \left(\mathbf{I}_T - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \right) \end{bmatrix} \mathbf{Y} \\ &= \begin{bmatrix} \left(\tilde{\mathbf{Z}}' \tilde{\mathbf{Z}} - \tilde{\mathbf{Z}}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \tilde{\mathbf{Z}} \right)^{-1} \tilde{\mathbf{Z}}' \left(\mathbf{I}_T - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \right) \end{bmatrix} \mathbf{Y} \\ &= \begin{bmatrix} \left(\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}} \right)^{-1} \tilde{\mathbf{Z}}' \mathbf{M}_X \end{bmatrix} \mathbf{Y} \\ &= \begin{bmatrix} \left(\boldsymbol{\beta}' \mathbf{Z}' \mathbf{M}_X \mathbf{Z} \boldsymbol{\beta} \right)^{-1} \boldsymbol{\beta}' \mathbf{Z}' \mathbf{M}_X \end{bmatrix} \mathbf{Y}. \end{aligned} \quad (23)$$

The second moment of \mathbf{B} is

$$\begin{aligned} \text{Var}(\text{vec}(\mathbf{B} | \boldsymbol{\beta}, \text{Data})) &= \frac{1}{T - k - r - n - 1} \tilde{\mathbf{S}} \otimes (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \\ &= \frac{1}{T - k - r - n - 1} (\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}} \mathbf{Y}) \otimes (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1}, \end{aligned} \quad (24)$$

¹⁰It is a more detailed exposition of the proof that can be found in Bauwens *et al.* (1999).

provided that $T - k - r > n + 1$.

4 Empirical results 1: Parametric and bootstrap procedures

In this Section we apply parametric and bootstrap techniques to the economic problem of inquiring whether and to what extent fiscal sustainability has been achieved in Poland. Subsection 4.1 deals with data issues. In Subsection 4.2 we perform preliminary unrestricted VAR estimation. In Subsection 4.3 we report unit root tests results, using parametric and bootstrap procedures. In Subsection 4.4 we apply the classical cointegration methodology (Engle and Granger, 1987 and Johansen, 1996) in order to determine the cointegration rank.

4.1 Data description

The variables used in the empirical part are total revenues (and grants) and total expenditures (inclusive of interest payments) of the consolidated central level for Poland. The source of the data is the IMF's *International Financial Statistics: Government Finance* (IFS2_8) for total revenues and grants (series 964:81) and total expenditures (series 964:82). Data refer to the main components of the central level, i.e. State budget, Social insurance fund, Pension and disability fund and Labour fund. The data frequency is monthly. Monthly data are lacking for 1996 and only quarterly statistics are presented for this year. Therefore, we start our analysis from January 1997. Data are seasonally unadjusted and are expressed in Polish currency (*zloty*). For each series, observations ranging from January 1997 to December 2006 are available¹¹.

Nominal revenues and expenditures must be deflated to work in real terms. We rely on price indexes¹², on a monthly basis, made available from ECB's *Indices of Consumer Prices* (ECB_ICP1). We choose as a measure of inflation the Polish harmonised index of consumer prices (HICP), overall index¹³. Data are seasonally unadjusted and unit index base is 2005=100.

To start, we proceed with a descriptive analysis. Note that we use the logarithmic transformation throughout the empirical part. Table 1 shows some descriptive statistics for log of deflated total revenues and log of deflated total expenditures.

[TABLE 1 ABOUT HERE]

We observe that all the variables have a strong dynamic component and are autocorrelated (see ρ_1 and ρ_2). These variables are not only correlated with their immediate past, but also with the observations of the same month of previous years (see ρ_{12} and ρ_{24}). This tendency is also confirmed by Ljung-Box Q-statistics.

In Figure 1, the top panel shows time series plots for log of deflated revenues and expenditures. As it can readily be seen upon inspection of this graph, the variables are not trending upwards and present

¹¹For more explanations, consult the webpage <http://www.mf.gov.pl/>. We do not include available data for 2007 since we prefer not to work with first release data, although we are well aware of the fact that macroeconomic data are subject to an indefinite number of revisions.

¹²We would have rather to use the GDP deflator, but this series is unavailable at the monthly frequency.

¹³Another reason to start in January 1997 is to avoid severe inflationary episodes which took place before this date, after the abolition of price controls.

systematic seasonal patterns. This feature is particularly evident for revenues. This is normal, since data provided by the authorities are not seasonally adjusted and, on the revenue side, largely reflect the calendar of payments. As a consequence, in the empirical analysis described in what follows, deterministic seasonality is given a special treatment. VAR estimation, stationarity tests and Bayesian cointegration analysis are all conducted including a set of deterministic seasonal dummies in the test regressions.

The bottom panel displays a cross-plot of log of deflated revenues and expenditures, indicating that both series share a possible long-run common equilibrium relationship.

[FIGURE 1 ABOUT HERE]

Looking at the top panel we notice an anomalous behaviour of revenues and expenditures in the second half of 1999, i.e. a sudden decrease followed by a sharp increase at the very end of 1999. A dummy variable (DUM_t) has been included in the data set to remove the influence of this portion (i.e. second half of 1999) of the data: This dummy variable is defined as $DUM_t = 1$ if t is equal to July 1999, August 1999, September 1999, October 1999, November 1999, December 1999, January 2000 and $DUM_t = 0$ elsewhere.

From a careful look, moreover, we can see that expenditures far exceed revenues, with very few exceptions. This gap is minimum at the beginning of the sample period and increases steadily after the end of year 2000. It seems slightly to reduce only at the end of 2006.

This tendency is also confirmed by Figure 2, which presents the consolidated gross debt (based on ESA 1995 definition), as a percentage of GDP at market prices, over the period 1996-2006. Note that the displayed data pertain to the General Government sector, which comprises the central level. The data source is the ECB's *AMECO* (ECB_AME1). There is a downward trend up to 2000, followed by an upward trend, indicating a clear deterioration of budget deficit. At the end of 2006, the debt ratio is approaching 50% of GDP.

[FIGURE 2 ABOUT HERE]

Concerns over a mounting debt, therefore, legitimately raise the issue of fiscal sustainability. In this framework, the aim of the empirical analysis is to assess whether the Polish government can continue to operate under its current fiscal policy indefinitely, using classical cointegration analysis to determine the cointegration rank and Bayesian inference to draw the posterior distribution of the unrestricted element of the cointegrating vector.

4.2 VAR estimation

As a *first* step, we estimate the unrestricted VAR model in (2) on revenues and expenditures. The VAR system is bivariate since $\mathbf{y}_t = \begin{pmatrix} rev_t \\ exp_t \end{pmatrix}$, where as already explained rev_t represents total revenues and exp_t total expenditures at time t . The estimated system is

$$\begin{bmatrix} rev_t \\ exp_t \end{bmatrix} = \begin{bmatrix} DET_{rev} \\ DET_{exp} \end{bmatrix} + \mathbf{A}_1 \begin{bmatrix} rev_{t-1} \\ exp_{t-1} \end{bmatrix} + \dots + \mathbf{A}_p \begin{bmatrix} rev_{t-p} \\ exp_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{rev} \\ \varepsilon_t^{exp} \end{bmatrix}, \quad (25)$$

with the deterministic part defined as

$$\begin{aligned} DET_{rev} &= a_{rev} + d_{rev}DUM_t + \sum_{s=2}^{12} \kappa_s^{rev} \delta_{s,t} \\ DET_{exp} &= a_{exp} + d_{exp}DUM_t + \sum_{s=2}^{12} \kappa_s^{exp} \delta_{s,t}, \quad t = 1, \dots, T, \end{aligned} \quad (26)$$

where DUM_t is a dummy variable that removes the influence of the anomalous time series behaviour in the second half of 1999 and $\delta_{s,t}$ ($s = 2, \dots, 12$) are a set of monthly deterministic seasonal dummies to be selected. These latter are expected to capture seasonal effects.

The order of the autoregressive model has to be chosen, as well. This selection is of crucial importance since it affects the outcome of the cointegration rank test that is going to be implemented. The lag length is determined by focusing on the residuals structure and by using lag exclusion tests. The optimum lag length of the VAR system is two ($p = 2$). The model is estimated in levels by means of OLS.

[TABLES 2 AND 3 ABOUT HERE]

The top part of Table 2 displays the unrestricted VAR estimated coefficients, the standard errors and the t-statistics. Note that the estimated model includes a constant term, a dummy variable for the second half of 1999 and a seasonal dummy for $s = 12$. Different deterministic seasonal dummies have been omitted because they are statistically not significant at 5%. This specification has been preferred since it delivers the highest adjusted R-squared, as a measure of goodness of fit, and the lowest value of the Schwarz criterion. These summary regression statistics are presented in the central part of Table 2.

In the bottom part of Table 2 it is shown the outcome of two VAR residual based tests. The residual serial correlation Lagrange Multiplier (LM) test is calculated for a lag order equal to 2. Under the null hypothesis of no serial correlation, the LM statistic is asymptotically distributed as a chi-square with 4 degrees of freedom. Lagrange Multiplier test statistic excludes serial correlation in the residuals at 5%. The residual Jarque-Bera normality test, under the null hypothesis (i.e. residuals are multivariate normal), is asymptotically distributed as a chi-square with 4 degrees of freedom. We fail to reject the null hypothesis of normality at 5% significance level. Finally, Table 3 reports the inverted roots of the characteristic AR polynomial. No root lies outside the unit circle, hence the estimated VAR system satisfies the stationarity condition¹⁴.

4.3 Parametric and bootstrap unit root tests

Fiscal sustainability is a long-run concept. To identify long-run equilibrium relationships between revenues and expenditures, graphical analysis must be supplemented by statistical inference. Specifically, cointegration tests may only be performed on series that are known to be non-stationary. Thus, in this Subsection, we implement single-equation based tests to examine whether revenues and expenditures contain zero frequency unit roots in their data generating process.

As a *second* step, therefore, we run an augmented Dickey-Fuller test (Said and Dickey, 1984), a commonly used unit root test. The null hypothesis is non-stationarity, i.e. presence of a zero frequency unit root. The augmented Dickey-Fuller (ADF) test corrects for higher order serial correlation by adding

¹⁴If the autoregressive polynomial is defined as the matrix polynomial in L (where L is the usual lag operator), i.e. $A(L) = I - A_1L - A_2L^2 - \dots - A_pL^p$, it is well known that the time series is stationary if the zeros of the determinantal polynomial $|A(z)|$ are outside the unit circle i.e. z^{-1} in $|A(z)| = 0$ equation. Expressed differently, the estimated VAR is stationary if all the roots have modulus less than one.

lagged differenced terms on the right-hand side of the test regression. To select the appropriate number of lagged first difference terms k to add, we make use of the recursive procedure proposed by Ng and Perron (1995). We start from an hypothetical maximum lag length, $k_{max} = 12$. The procedure requires to test downward, reducing k by one until the last lag becomes statistically significant (at 5% level). If no lags are found to be significant, then the selected lag length is zero. In this way we expect to remove any serial correlation in the residuals.

For ADF test, some preliminary regressions have been tried with 1) an intercept and 2) with an intercept and a linear time trend. The time trend is not found to be significant¹⁵, at 5% significance level. However, some care should be taken regarding the inclusion of an intercept in the ADF test. For instance Buseti *et al.* (2006) point out that, to enhance their power, unit root tests should be applied without allowing for an intercept term in the test regression.

Besides, as already highlighted, we are working with data that exhibit strong seasonality. To deal with this problem, several alternatives are at hand: First, seasonal unit root tests (e.g. Hylleberg *et al.*, 1990, HEGY) may be applied. Second, a possibility is to work with filtered data, in order to eliminate the eleven seasonal unit roots, while keeping the zero frequency unit root. Third, if the seasonality is regular, it is easy to perform unit root tests with deterministic seasonal dummies¹⁶. This latter is the route we take.

As shown by Demetrescu and Hassler (2007) via Monte Carlo simulations, neglecting deterministic seasonality may distort the asymptotic distribution of the ADF test statistics. In particular, when deterministic seasonality can not be *a priori* excluded, the ADF test run with only the constant and without seasonal dummies may suffer from severe size distortions. The inclusion of seasonal dummies, useful to discern deterministic seasonal effects, is highly recommended. For this reason we specify the single-equation test regression adding, in the deterministic part, deterministic seasonal dummies (together with DUM_t , previously defined),

$$\begin{aligned}\Delta y_t &= d_y DUM_t + \sum_{s=1}^{12} \kappa_s^y \delta_{s,t} + (\rho - 1)y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + u_t, \\ y_t &= rev_t \vee y_t = exp_t, \quad t = k + 2, \dots, T,\end{aligned}\tag{27}$$

where $\delta_{s,t} = 1, 0$ depending on whether t falls in month s or not.

[TABLE 4 ABOUT HERE]

Table 4 reports the outcomes of the ADF tests for the series in levels, based on (27). As explained by Dickey *et al.* (1986), the inclusion of seasonal dummies does not affect the asymptotic distribution of the Dickey-Fuller test. The results do not imply rejection of the presence of a unit root for both series in levels. This holds true for both series, however, only at 1% significance level. At 5%, indeed, the null hypothesis is rejected for revenues. In summary, focusing on the ADF t-statistics and on MacKinnon (1996) critical values, results are clear-cut for expenditures (i.e. empirical evidence suggests that they are $I(1)$ variables) while they are more controversial for revenues.

However, as noted by Davidson and Flachaire (2001), among others, another potential problem of unit root tests is that they can be severely biased in small samples. To correct for this drawback and

¹⁵In addition, the zero mean hypothesis for revenues and expenditures series (both taken in first differences) is tested. We fail to reject the null hypothesis of zero mean for both series, at 5% and 1% significance level. Therefore we conclude that $\mathbb{E}(\Delta y_t) = 0$, where y_t are revenues or expenditures, implying no growth in the variables. Results are available from the author upon request.

¹⁶A not very attractive alternative is to work with seasonally adjusted data, since it has been found that seasonal filters may affect the conclusion of the test. To avoid this drawback, we exclude this possibility.

to check the robustness of our results, we resort to bootstrap methodology. Bootstrap methods rely on simulation to approximate the finite sample distribution of the test statistic under the null hypothesis. As explained by Chang *et al.* (2006), if properly implemented, bootstrap resampling schemes provide better approximation to the finite sample distributions of the statistics than their asymptotic counterparts.

We apply the bootstrap to obtain appropriate 5% critical values of the distribution of the ADF t-test statistics under the null hypothesis of non-stationarity. We do this following the *sieve* scheme. We refer to Psaradakis (2001) and Chang and Park (2003), which are theoretical papers dealing with sieve bootstrap ADF and DF tests. These authors establish the asymptotic validity of sieve bootstrap unit root tests under general conditions. We construct the sieve bootstrap distribution following the algorithm described by Chang and Park (2003) and summarized by Palm *et al.* (2006). In doing this, the practical issues on the sieve bootstrap procedure highlighted in Chang and Park (2003), p.3, have been taken into careful consideration¹⁷.

Figure 3 presents the sieve bootstrap distributions for ADF test statistics computed for revenues and expenditures, based on equation (27). The number of replications in the Monte Carlo experiment is 5,000.

[FIGURE 3 ABOUT HERE]

As it is evident from a careful look at the bottom part of Table 4, the 5% critical values obtained from the sieve bootstrap distribution of the ADF t-statistics help to shed light on our results: In general, in the case under study, the bootstrapped 5% critical values are rather different with respect to those reported by MacKinnon (1996). In particular, using sieve bootstrap critical values, the null hypothesis of non-stationarity can not be rejected at the reported significance level. This finding enables us to proceed with cointegration analysis, testing the existence of a cointegrating relationship.

4.4 Cointegration rank analysis

Hereafter we assess fiscal sustainability identifying the long-run equilibrium relationship between revenues and expenditures. As explained by Quintos (1995), fiscal policy is *strongly* sustainable if the cointegrating vector between revenues and expenditures is equal to $[-1 \ 1]'$. As a consequence, the strong sustainability condition requires to test the null hypothesis that these two series are cointegrated and that the cointegrating vector is $[-1 \ 1]'$. This is done running the cointegration test (Johansen, 1996), whose purpose is to determine whether a group of non-stationary series are cointegrated. As already detailed, an impact matrix $\Pi = -A(1)$ in (3) of reduced rank, i.e. smaller than its dimension, is a feature of cointegrated VAR systems. In a nutshell, Johansen's methodology requires to estimate the Π matrix and to test, at the same time, whether the restrictions implied by the reduced rank of Π can be rejected or not.

Graphical evidence seems to suggest that revenues and expenditures level data have a non-zero constant and no deterministic trends. Whether this feature applies to an eventual cointegrating relationship between them is not easy to discern by graphical inspection, although a deterministic trend does not seem to be present. On the contrary, there are doubts regarding a non-zero intercept inside the cointegration space. This is worth pointing out since, in general, the asymptotic distribution of the rank test statistic for cointegration is not standard and strictly depends on the assumptions made with respect to the deterministic components of the VECM model in (3). Only when the model specification is defined, it is possible to determine the rank of the long-run multiplier Π matrix.

To select among possible models, we apply the log-likelihood ratio test, i.e. a test in which the maximum of the likelihood function of a restricted model is compared to the maximum of the likelihood

¹⁷The developed MATLAB code is available from the author upon request.

function with the constraints relaxed. In particular, we estimate two different VECM specifications¹⁸: 1) according to the first VECM model, the level data do not possess a deterministic trend and the cointegrating equation does not have an intercept; 2) according to the second VECM model, the level data do not possess a deterministic trend but the cointegrating equation has an intercept. These, in our view, are the most plausible alternatives. That is, the model in 1) differs from the model in 2) only by the addition of one parameter, the constant term in the cointegration space¹⁹. We then calculate the test statistic $\lambda = -2(\log(L_0) - \log(L_1))$, where $\log(L_0)$ is the log-likelihood of the restricted model (intercept set equal to zero as specified in the null hypothesis), whereas $\log(L_1)$ is the log-likelihood of the unrestricted model. Under the null hypothesis, λ is distributed as a chi-square random variable with 1 degree of freedom. We fail to reject the null hypothesis at 5% significance level. We therefore favour specification 1), implying no growth in the level data and zero intercept in the cointegrating relation²⁰.

On this basis, the presence of a cointegrating relationship is tested by using the trace test. At this stage, therefore, Johansen's *maximum likelihood based* methodology is used. We rely on this strategy since, as explained by Bauwens *et al.* (1999), the selection of the cointegration rank applying Bayesian inference has not yet given very convincing results.

[TABLE 5 ABOUT HERE]

The outcome of the trace test is presented in the top part of Table 5. The null hypothesis is no cointegrating vectors among the variables. We do not include an intercept in the cointegrating equation. Deterministic trends are excluded, as well. The critical values are those reported by MacKinnon *et al.* (1999). At the 5% significance level, this test rejects the presence of no cointegration among the series, in favour of a single cointegrating relationship ($r = 1$)²¹. This suggests that revenues and expenditures share one cointegrating relationship and there is one common stochastic trend, i.e. $\mathbf{y}_t \sim CI(1,1)$. There exists a linear combination of these two difference-stationary series so that the residuals obtained from this linear combination are stationary.

The restricted coefficients derived from the cointegrating equation are displayed in the bottom part of Table 5. These coefficients are the Johansen's maximum likelihood (ML) estimates of the exactly identified cointegrating vector, after normalization. The estimated cointegrating vector is $\beta = [-1 \ 0.8638]'$. The associated equilibrium equation $\beta' \mathbf{y}_t = 0$ corresponds²² to $-rev_t + 0.8638exp_t = 0$. The asymptotic standard error of the sample statistic is 0.1262.

On balance, revenues seem to grow slower than expenditures. The robustness of this result, in any case, should be carefully checked. We do this in the sequel, deepening our analysis and applying Bayesian inference to the cointegrated VAR model.

5 Empirical results 2: Marginal posterior densities, prior and posterior parameters

In Section 4 we have determined the number of cointegrating relationships by means of the trace test. We find one cointegrating equation between revenues and expenditures in the estimated bivariate VAR

¹⁸Estimation results are available from the author upon request.

¹⁹Note that, therefore, we use the likelihood ratio test to compare two hierarchically nested models.

²⁰We note that, without restricting the constant term to zero in the cointegration equation, it would suggest that in the long-run revenues can be positive even if expenditures are zero.

²¹We point out that, since we are estimating a 2×2 VAR system, maximum-eigenvalue and trace tests should produce identical results if $r = 1$. For this reason the maximum-eigenvalue test statistic is not reported.

²²Since we are dealing with logs of deflated revenues and expenditures, we should better write $-\log(rev_t) + \beta_{exp} \log(exp_t) = 0$. In what follows we shall stick to this notation.

system. Therefore, the rank of cointegration is equal to one ($r = 1$) and β is a 2×1 cointegrating vector.

As already discussed, the cointegrating vector is not identified unless we introduce some arbitrary normalization. In general, to be identified, r^2 restrictions have to be imposed on α and β matrices. Since in our case $r = 1$, there is a single normalization for the cointegrating vector, that corresponds to $\beta = [-1 \ \beta_{exp}]'$, normalizing on revenues. Moreover, α' is a 1×2 vector of adjustment parameters. Differently from β , as already explained, it is left unrestricted.

In what follows, we focus on the case of a single cointegrating vector. The setup is the VAR model in (2) that, if cointegration holds, may be written in error correction form as in (4). This latter corresponds to the regression model in (6). This representation is needed to perform the Bayesian empirical analysis. Note that, henceforth, we assume for the vectors D_0, D_1, \dots, D_{T-1} in (7) the same deterministic structure reported in (26).

Sticking to the regression model in (8) and using the prior in (11), Bauwens *et al.* (1999) provide some useful results concerning the analytical form of the posterior density of the cointegrating vector β . In Subsection 5.1 we review part of these issues. In Subsection 5.2 we deal with elicitation of prior parameters, while in Subsection 5.3 we briefly explain the methodology used to calculate moments of the marginal posterior distribution for β . In Subsection 5.4 we present and discuss estimation results.

5.1 Posterior analytical results

We are dealing with a single cointegrating vector. Any prior density on the unrestricted element of β , $\phi(\beta_{exp})$, can be used, at least in principle. If we assume as prior density either a flat $\phi(\beta_{exp}) \propto 1$ or a Student distribution, the posterior density is integrable and belongs to a known class, the poly-t densities. As explained by Drèze (1977), no analytical expressions exist for moments of these densities. Hence, we have to resort to numerical integration. Since we compute posterior distributions applying simulation methods, there is no reason to restrict our analysis to Student or flat priors. Other prior distributions may be used, as well.

Let us denote with s the number of linear restrictions imposed on β in addition to the normalization condition: s is known as the order of over-identification. We report without proving the following results, that are Corollary 9.4 in Bauwens *et al.* (1999)²³:

1. If the prior on the unrestricted elements of β is diffuse, i.e. $\phi(\beta_{exp}) \propto 1$, the posterior $\phi(\beta_{exp}|Data)$ is a 1-1 poly-t integrable density. If, additionally, the order of over-identification is null ($s = 0$), this density does not possess finite moments.
2. If, on the contrary, $\phi(\beta_{exp})$ is Student with ν_0 degrees of freedom, the posterior $\phi(\beta_{exp}|Data)$ is a 2-1 poly-t integrable density with finite moments of order $n + \nu_0 - 1$.

Since we impose no additional restrictions on the β vector, $s = 0$ in the case under study. Therefore, with a diffuse prior, the 1-1 poly-t density has no finite moments²⁴. To avoid this drawback, let us assume a Student prior density on β_{exp} with ν_0 degrees of freedom. The posterior density in (20) corresponds to a 2-1 poly-t that can be specified as

$$\phi(\beta_{exp} | Data) \propto [1 + m_0(\beta_{exp} - \beta_0)^2]^{-\frac{1}{2}(\nu_0+1)} \times \frac{|\beta_{exp}^2 \mathbf{W}_2|^{l_2}}{|\beta_{exp}^2 \mathbf{W}_1|^{l_1}}, \quad (28)$$

where β_0 , m_0 , ν_0 are prior parameters and $[1 + m_0(\beta_{exp} - \beta_0)^2]^{-\frac{1}{2}(\nu_0+1)}$ is the corresponding Student kernel.

²³For the proof we refer to this book.

²⁴However, defining a truncated flat prior, finite moments do exist.

First, we focus on the ratio of determinants on the right-hand side of (28). Remind from (21) that $\mathbf{W}_2 = \mathbf{Z}'\mathbf{M}_X\mathbf{Z}$ and $\mathbf{W}_1 = \mathbf{Z}'\mathbf{M}_Y(\mathbf{I}_T - \mathbf{Y}(\mathbf{Y}'\mathbf{M}_Y\mathbf{Y})^{-1}\mathbf{Y}')\mathbf{M}_Y\mathbf{Z}$. These 2×2 matrices may be represented as $\mathbf{W}_2 = \begin{bmatrix} W_2^{(1,1)} & W_2^{(1,2)} \\ W_2^{(2,1)} & W_2^{(2,2)} \end{bmatrix}$ and $\mathbf{W}_1 = \begin{bmatrix} W_1^{(1,1)} & W_1^{(1,2)} \\ W_1^{(2,1)} & W_1^{(2,2)} \end{bmatrix}$, where with $\mathbf{W}_k^{(i,j)}$ we indicate the (i, j) -th entry of the k -th \mathbf{W} matrix.

Aside from constant terms, it can be shown that the analytical form of the posterior may be further expressed as

$$\phi(\beta_{exp} \mid Data) \propto [1 + m_0(\beta_{exp} - \beta_0)^2]^{-\frac{1}{2}(\nu_0+1)} \times \frac{(s_2 + m_2(\beta_{exp} - \beta_2)^2)^{l_2}}{(s_1 + m_1(\beta_{exp} - \beta_1)^2)^{l_1}}, \quad (29)$$

where $m_2 = W_2^{(2,2)}$, $m_1 = W_1^{(2,2)}$, $\beta_2 = m_2^{-1}W_2^{(1,2)}$, $\beta_1 = m_1^{-1}W_1^{(1,2)}$, $s_2 = W_2^{(1,1)} - \beta_2^2 m_2$ and $s_1 = W_1^{(1,1)} - \beta_1^2 m_1$. Note that poly-t distributions are characterized by the property that their density kernels are products, or products of ratios, of Student density kernels, exactly as in (29). We refer to Dr ze (1977) for a review of poly-t densities in the framework of Bayesian regression analysis.

Of course, other prior densities can be introduced in (28). For instance, a suitable choice seems also to be

$$\phi(\beta_{exp} \mid Data) \propto [\beta_{exp}^{\iota-1}(1 - \beta_{exp})^{\kappa-1}] \times \frac{(s_2 + m_2(\beta_{exp} - \beta_2)^2)^{l_2}}{(s_1 + m_1(\beta_{exp} - \beta_1)^2)^{l_1}}, \quad (30)$$

where $\iota > 0$ and $\kappa > 0$ are prior parameters and $[\beta_{exp}^{\iota-1}(1 - \beta_{exp})^{\kappa-1}]$ is the corresponding kernel of a Beta density, i.e. $\beta_{exp} \sim \text{Beta}(\iota, \kappa)$ with $0 < \beta_{exp} < 1$. Note, however, that with a Beta prior the parameter domain is constrained on the unit simplex. However, the Beta density on $(0,1)$ can be easily transformed to a Beta density on any finite interval by a linear change of variable.

5.2 Elicitation of prior parameters

Let us now turn to the specification of sensible prior distributions and parameters, whose elicitation is perhaps the less obvious task in implementing Bayesian inference. In (29), the Student prior mean may reasonably be $\beta_0 = 1$. This is set on the grounds that, at least in the long-run, we expect revenues and expenditures to grow at the same rate and fiscal policy to be sustainable. If the prior density is specified by means of a Student density centred at one, where is the posterior density $\phi(\beta_{exp} \mid Data)$ located? This is the kind of question we try to answer herein.

To elicit m_0 , we can solve for the unknown parameter the equation $m_0 = \frac{1}{\nu_0-2}\text{Var}(\beta_{exp})^{-1}$, if we have prior information on $\text{Var}(\beta_{exp})$. Little indication is given by the data concerning the prior variance. One possibility is to solve this equation for different values of $\text{Var}(\beta_{exp})$. In this way we can assess how the prior dispersion influences the posterior density and its parameters. Alternatively, we can follow the method described by Bauwens *et al.* (1999), p.103, for the elicitation of prior parameters of marginal Student densities.

According to this latter approach, we set $s_0 = 1$, $\nu_0 = 12$, $\beta_0 = \frac{\bar{q}_{0.025} + \bar{q}_{0.975}}{2} = \frac{0.5+1.5}{2} = 1$, $m_0 = \frac{1}{\nu_0} \left(\frac{2t_{0.95}}{\bar{q}_{0.975} - \bar{q}_{0.025}} \right)^2 = 1.0589$, where $t_{0.95} = 1.7823$ is the 95 percentile of the standardized Student density with 12 degrees of freedom. Consequently, $\text{Var}(\beta_{exp}) = 0.0944$. Note that we have chosen for β_{exp} the 95 per cent highest prior density interval $Pr[\bar{q}_{0.025} < \beta_{exp} < \bar{q}_{0.975}] = Pr[0.5 < \beta_{exp} < 1.5] = 0.95$. A word of caution at this point: In Subsection 5.4, we are going to *truncate* the Student (and Beta) priors to $(-1, 3)$. The moments $\beta_0 = 1$ and $\text{Var}(\beta_{exp}) = 0.0944$ are not exact for *truncated* Student (and Beta) priors. However, these values are close to the exact ones since the truncation is 'small' (see Figures 4 and 5).

[FIGURE 4 ABOUT HERE]

Figure 4 displays the Student ($\beta_0 = 1, s_0 = 1, m_0 = 1.0589, \nu_0 = 12$) prior, the marginal likelihood and corresponding posterior density of the sampled data. These are our reference values. The posterior density is slightly right-skewed, reflecting the influence of the likelihood of the data. Tails are not so long since, in the case under study, the posterior density is integrable and possesses finite moments. As we may expect, with this choice of the β_0 prior parameter, the prior information on β_{exp} moves the posterior expected value towards the prior expected value. Clearly enough, moreover, the posterior density is located between the likelihood and the prior density. All this is visible, *ictu oculi*, in Figure 4.

Furthermore, the smaller the m_0 parameter, the bigger is the prior variance $\text{Var}(\beta_{exp})$. The bigger is $\text{Var}(\beta_{exp})$, the less the prior is concentrated around its mean. Thus, also the posterior density is influenced by the choice of m_0 prior parameter.

Focusing on the Beta prior density in (30), we initially suggest a Beta(2, 2) specification ($\iota = \kappa = 2$), i.e. a symmetric distribution around the mean ($\mathbb{E}(\beta_{exp}) = \frac{\iota}{\iota + \kappa} = 0.5$). However, a prior distribution on (0,1) is excessively restrictive as hypothesis. Therefore, we apply a linear change of variable to β_{exp} , generating a new variable $\bar{\beta}_{exp} = c \times \beta_{exp} + d$, to be defined in a larger interval. For instance, taking $c = 3.5$ and $d = -1$, we have that $\mathbb{E}(\bar{\beta}_{exp}) = c \times \mathbb{E}(\beta_{exp}) + d = 0.75$. Alternatively, if $c = 4$ and $d = -1$, $\mathbb{E}(\bar{\beta}_{exp}) = c \times \mathbb{E}(\beta_{exp}) + d = 1$. This last choice for c and d constants is implemented hereafter.

Coming to the prior variance, $\text{Var}(\beta_{exp}) = \frac{\iota\kappa}{(\iota + \kappa + 1)(\iota + \kappa)^2} = 0.05$. If $c = 3.5$ and $d = -1$, $\text{Var}(\bar{\beta}_{exp}) = c^2 \times \text{Var}(\beta_{exp}) = 0.6125$, being the variance a quadratic operator. Similarly, if $c = 4$ and $d = -1$, $\text{Var}(\bar{\beta}_{exp}) = c^2 \times \text{Var}(\beta_{exp}) = 0.8$. The likelihood dominates the Beta prior much more than the Student prior, since the Beta is more dispersed than the Student. We note that the Beta(2, 2) prior variance is almost nine times bigger than Student (1, 1, 1.0589, 12) prior variance (0.0944). The Beta(2, 2) density is too dispersed and, therefore, the posterior results are not fully comparable with those obtained assuming a Student prior. For this reason, we change the parameters of the Beta prior to make its prior variance closer to that of the Student. We keep the Beta defined on the (-1, 3) interval and we increase the two parameters ι and κ to decrease the prior variance. Moreover, to ensure a symmetric shape of the distribution, we require $\iota = \kappa$.

By using the explicit formula for the variance of a Beta prior and the fact that ι is equal to κ , we find that

$$\text{Var}(\bar{\beta}_{exp}) = c^2 \times \text{Var}(\beta_{exp}) = 16 \times \left(\frac{\iota\kappa}{(\iota + \kappa + 1)(\iota + \kappa)^2} \right) = 0.0944 \Rightarrow \iota = \kappa = 20.6864.$$

Consequently, we choose a Beta($\iota = 20, \kappa = 20$), truncated on (-1, 3), as the final prior specification.

[FIGURE 5 ABOUT HERE]

Figure 5 displays a Beta($\iota = 20, \kappa = 20$) prior, the marginal likelihood and the corresponding posterior density. The Beta prior is centred at 1, exactly as the Student prior in Figure 4. The Beta prior variance is $\text{Var}(\bar{\beta}_{exp}) = 0.0976$. The prior domain is (-1, 3).

5.3 Posterior estimation methods

Moments of the posterior densities in (29) and (30) are not known analytically. Numerical integration has to be used to obtain them²⁵. In principle, one way to compute posterior results on parameters is by posterior simulation methods. Bauwens and Giot (1998) suggest to apply Markov chain Monte Carlo methods (specifically, Gibbs sampling and Griddy-Gibbs sampling techniques) to a cointegrated VAR

²⁵Bauwens and Richard (1985) propose an algorithm to simulate poly-t densities.

system of aggregate imports and import prices for Belgium. Another possibility is to use importance sampling (e.g. Bauwens and Lubrano, 1996). This sampling scheme requires to choose an importance function that can be simulated directly as a good approximation to the poly-t posterior density to evaluate. This choice, generally speaking, is not an easy task to accomplish.

Another way to calculate posterior results on parameters, especially when the dimension of the integration is not large, is to make use of deterministic integration. Tompa (1973) points out that deterministic integration is very efficient for computing the integral, the moments and the marginal densities of m-1 and m-0 poly-t densities in k variables (whatever the value of m and k). Moreover, this method is useful and easy to implement for computing an integral of dimension one, as in the case under study. Therefore, in the sequel, we are going to use deterministic integration to derive the posterior distribution and to evaluate the first two posterior moments.

Focusing on computational issues, posterior results using deterministic integration are obtained by implementing the extended Simpson's method. It consists in interpolating the integrand by a generic polynomial that matches the integrand at some selected points. Let $\phi(\beta_{exp})$ be the function to be approximated and β_{exp} a scalar. We assume that the integration range is $(0, 1)$. For different ranges²⁶, it is always possible to apply a one-to-one transformation of variable to calculate the integral over $(0, 1)$. The function is approximated through $2n + 1$ equally spaced points, according to the relation

$$\int_0^1 \phi(\beta_{exp}) d\beta_{exp} \cong \frac{1}{6n} \left[\phi(\beta_{exp}^{(0)}) + 4\phi(\beta_{exp}^{(1)}) + 2\phi(\beta_{exp}^{(2)}) + \dots + 2\phi(\beta_{exp}^{(2n-2)}) + 4\phi(\beta_{exp}^{(2n-1)}) + \phi(\beta_{exp}^{(2n)}) \right],$$

where on the right-hand side we find the interpolating polynomial at $2n + 1$ data points. Obviously, the larger the number of points, the better the quality of the approximation²⁷.

Note that, as explained in Bauwens and Lubrano (1996), at the same time numerical posterior results on β_{exp} are calculated, it is possible to obtain estimates of α vector. From (17), indeed, we already know that $\mathbf{B} | \beta, Data \sim Mt_{(m+n(p-1)+r) \times n}(\tilde{\mathbf{B}}, \tilde{\mathbf{W}}' \tilde{\mathbf{W}}, \tilde{\mathbf{S}}, T - (m + n(p-1)) - r)$, i.e. \mathbf{B} (and hence α') follows a matrixvariate Student distribution, conditionally on β . Moreover, each element of α' follows a univariate Student distribution.

The first two conditional moments of α vector (given β) are reported in (23) and (24). At each point the posterior $\phi(\beta | Data)$ is evaluated, the posterior $\phi(\alpha' | \beta, Data)$ is also evaluated. In this way, posterior results on α' may be obtained simply marginalizing the conditional results on α' given β .

5.4 Posterior estimation results and interpretation

We apply deterministic integration to calculate posterior results for $\phi(\beta | Data)$ and $\phi(\alpha' | \beta, Data)$. Different choices concerning the prior densities and the most appropriate values of the parameters are made (we refer to Subsection 5.2 for a description of the elicitation of the prior density).

Table 6 presents posterior results (mean and standard deviation of β_{exp})²⁸ assuming

²⁶For instance, (a, b) , $(-\infty, b)$, $(a, +\infty)$ or $(-\infty, +\infty)$.

²⁷Another deterministic method is the trapezoidal rule. It requires to approximate the integrand $\phi(\beta_{exp})$ by a linear function through the selected points. With $2n + 1$ equally spaced points, it delivers the the integral as

$$\int_0^1 \phi(\beta_{exp}) d\beta_{exp} \cong \frac{1}{4n} \left[\phi(\beta_{exp}^{(0)}) + 2\phi(\beta_{exp}^{(1)}) + 2\phi(\beta_{exp}^{(2)}) + \dots + 2\phi(\beta_{exp}^{(2n-2)}) + 2\phi(\beta_{exp}^{(2n-1)}) + \phi(\beta_{exp}^{(2n)}) \right].$$

In general, Simpson's method and the trapezoidal rule should be efficient only if the integrand function is very cheap to evaluate. Moreover, for given n , Simpson's rule should be more precise than the trapezoidal rule. For this reason, in what follows, we only present results obtained by applying Simpson's rule. We refer to Davis and Rabinowitz (1975) for further insights on deterministic methods.

²⁸The developed MATLAB code is available from the author upon request.

1. A Student($\beta_0 = 1, s_0 = 1, m_0 = 1.0589, \nu_0 = 12$) prior truncated on $(-1, 3)$ ²⁹;
2. A Beta($\iota = 20, \kappa = 20$) prior truncated on $(-1, 3)$;
3. A truncated flat prior on $(-1, 3)$.

Note that, in carrying out Bayesian analysis, $\mathbf{D}_0', \mathbf{D}_1', \dots, \mathbf{D}_{T-1}'$ comprised in the matrix \mathbf{X} in (7) contain a constant term, a deterministic seasonal dummy and a dummy variable to reduce the influence of the anomalous time series behaviour in the second half of 1999, as shown in (26).

[TABLE 6 ABOUT HERE]

As Table 6 reports, with a Student $(1, 1, 1.0589, 12)$ prior (case 1), the estimated posterior mean of β_{exp} is $\mathbb{E}(\beta_{exp}) = 0.9296$. The corresponding standard deviation is $\text{s.d.}(\beta_{exp}) = 0.2285$. We may wish to vary the Student prior parameters (ν_0 and m_0) to assess how they influence the estimates of the posterior moments³⁰. Although not shown in Table 6, calibrating a $(1, 1, 1.0589, 6)$ Student density, we get $\mathbb{E}(\beta_{exp}) = 0.9288$ and $\text{s.d.}(\beta_{exp}) = 0.2345$. Hence the influence seems to be limited, at least in this particular case. With a $(1, 1, 0.5, 12)$ Student prior, conversely, $\mathbb{E}(\beta_{exp}) = 0.9104$ and $\text{s.d.}(\beta_{exp}) = 0.2818$. This makes sense, of course: If the prior is less concentrated (i.e. m_0 is smaller), the posterior becomes less concentrated, as well.

We also present other distributional hypotheses concerning the prior density. Using a Beta(20, 20) (case 2) defined in $(-1, 3)$, the posterior mean is 0.9232 and the standard deviation increases to 0.2392. For a truncated flat prior (case 3), the estimated posterior mean is equal to 0.8936. The corresponding standard deviation is 0.4918. As expected, with a flat prior, posterior estimation is less accurate than when assuming informative priors (cases 1 and 2). Notice that the posterior density with a truncated flat prior corresponds to the marginal likelihood function in Figure 4 or in Figure 5 (dashed line). This density has positive skewness and fat tails.

Clearly enough, the posterior mean $\mathbb{E}(\beta_{exp})$ is located pretty far away from 1, whatever the prior density defined (cases 1,2,3), since the marginal likelihood of the cointegrating vector has more mass on the left of 1 and the specified priors are quickly dominated by the data. On balance, the estimates of the cointegrating coefficient for revenue-expenditure variables vary between 0.8936 and 0.9296. These estimates should be regarded as a measure of β_{exp} in the long-run equation $\log(rev_t) = \beta_{exp} \log(exp_t)$, where these variables have been expressed in real terms dividing by HICP. Interpreting these estimates in terms of elasticities, this indicates that, in the long-run, a 100 percent rise in budget expenditures will approximately translate into a 90 percent growth of revenues. And that Polish fiscal policy does not satisfy the intertemporal budget constraint (in strong form). Thus, on the basis of the available data, we have evidence that expenditures (inclusive of interest payments) are growing faster than revenues. Undoubtedly, this constitutes a warning signal to the Polish government. These results cast doubt on the possibility that government revenues are capable of meeting government expenditures, in equilibrium (in the long-run).

The bottom part of Table 6 presents the Bayesian posterior mode of the unrestricted element of the cointegrating vector (β_{exp}), assuming an untruncated flat prior. In this case the marginal posterior density is integrable but does not possess finite moments. The posterior mode is calculated by maximum likelihood, i.e. by maximizing with respect to β_{exp} the marginal likelihood function

$$\mathcal{L}(\beta_{exp} \mid Data) \propto \frac{(s_2 + m_2(\beta_{exp} - \beta_2)^2)^{l_2}}{(s_1 + m_1(\beta_{exp} - \beta_1)^2)^{l_1}}, \quad (31)$$

²⁹ Concerning the prior degrees of freedom, for large values of ν_0 , the Student density is close to the normal density. Conversely, for small values of ν_0 , it is characterized by fat tails.

³⁰ Additional results are available from the author upon request.

where, as explained in (21), $m_2 = W_2^{(2,2)}$, $m_1 = W_1^{(2,2)}$, $\beta_2 = m_2^{-1}W_2^{(1,2)}$, $\beta_1 = m_1^{-1}W_1^{(1,2)}$, $s_2 = W_2^{(1,1)} - \beta_2^2 m_2$ and $s_1 = W_1^{(1,1)} - \beta_1^2 m_1$. The maximum likelihood estimate and the associated asymptotic standard error have been obtained using the MATLAB function for optimization `Max_lik.m`³¹. The mode of the marginal likelihood function (i.e. the posterior density under a flat prior) is shifted to the left with respect to the posterior mean under a Student prior (0.9296) and under a Beta prior (0.9232): Indeed, as reported in Table 6, the mode of the marginal likelihood function is 0.8006 and the corresponding asymptotic standard error is 0.2883.

One comment about the Bayesian posterior mode (0.8006) and the Johansen's ML estimate (0.8638) reported in Table 5: The difference between them is not negligible and a few words of clarification are at order. As explained by Bauwens and Lubrano (1996), p.11, the Johansen's ML estimate corresponds to the maximum of the *concentrated* likelihood function, while the posterior mode under a flat prior is the maximum of the *marginal* likelihood function. In particular, it is possible to show that the concentrated likelihood function is

$$\tilde{\mathcal{L}}(\beta_{exp} | Data) \propto \frac{(s_2 + m_2(\beta_{exp} - \beta_2)^2)^{l_1}}{(s_1 + m_1(\beta_{exp} - \beta_1)^2)^{l_1}}, \quad (32)$$

i.e. it differs from the marginal likelihood function in (31) by the exponent in the numerator (l_1 instead of l_2).

[FIGURE 6 ABOUT HERE]

The densities in (31) and (32) are displayed in Figure 6. Note that both are 1-1 poly-t. It is evident that the concentrated likelihood function is shifted to the right with respect to the marginal likelihood function. Moreover, the concentrated likelihood function has fatter tails and, as expected, its maximum (0.8638) is located on the right of the maximum of the marginal likelihood function (0.8006).

Finally, let us turn to the dynamics of short-term adjustment towards the long-run equilibrium.

In Table 7 we present posterior results for the adjustment coefficients (i.e. α_{rev} and α_{exp}). In other words, focusing on the VECM equation in (4), we apply Bayesian inference to the estimation of α coefficients in the linear adjustment process $\alpha\beta' y_{t-1}$, which can be specified as

$$\begin{bmatrix} \alpha_{rev} \\ \alpha_{exp} \end{bmatrix} \begin{bmatrix} -1 & \beta_{exp} \end{bmatrix} \begin{bmatrix} rev_{t-1} \\ exp_{t-1} \end{bmatrix} = \begin{bmatrix} \alpha_{rev} \\ \alpha_{exp} \end{bmatrix} [-rev_{t-1} + \beta_{exp} exp_{t-1}]. \quad (33)$$

In (33), α_{rev} and α_{exp} are a measure of the speed of adjustment of rev_t and exp_t , respectively, towards their long-run equilibria. The higher their values, the quicker the adjustment process.

[TABLE 7 ABOUT HERE]

The estimation is performed bearing in mind the distributional result in (17), i.e. $\mathbf{B} | \beta, Data \sim Mt_{(m+n(p-1)+r) \times n}(\tilde{\mathbf{B}}, \tilde{\mathbf{W}}' \tilde{\mathbf{W}}, \tilde{\mathbf{S}}, T - (m + n(p-1)) - r)$. Expressions for $\mathbb{E}(\alpha' | \beta, Data)$ and $\mathbb{V}ar(\text{vec}(\alpha' | \beta, Data))$ are given in (23) and (24). Therefore, at each point the posterior density for β vector is evaluated, it is possible to calculate posterior conditional results for α . Subsequently, by marginalization, we can approximate $\mathbb{E}(\alpha' | Data)$ with

$$\sum_{i=1}^N \frac{\mathbb{E}(\alpha' | \beta_i, Data)}{N},$$

³¹Implemented by Michael Rockinger and available at <http://www.hec.unil.ch/matlabcodes/>.

where β_i is the marginal posterior density of β evaluated at the i -th data point ($i = 1, 2, \dots, N$), following the posterior estimation methods outlined in Subsection 5.3.

The estimated posterior mean of α_{rev} is $\mathbb{E}(\alpha_{rev}) = 0.3899$. The corresponding standard deviation is 0.3463. Moreover, the estimated posterior mean of α_{exp} is $\mathbb{E}(\alpha_{exp}) = -0.0561$, with a standard deviation of 0.1907. Expenditures are not 'correcting' in such a way to preserve the linear long-run relationship. Revenues, on the contrary, seem to adjust back towards equilibrium. Therefore, expenditures could be considered as weakly exogenous inside the cointegrating relationship. Not surprisingly, expenditures are driving revenues and determine the long-run equilibrium path over the observed sample period.

6 Concluding remarks

In this paper we have examined the sustainability of fiscal regime in Poland, inquiring whether, if pursued indefinitely, current policies would satisfy the intertemporal budget constraint. Our findings have to be treated cautiously, as the whole analysis is based on only ten years of monthly data, corresponding to the post-transition period. Such a data set is relatively short in the context of the sustainability debate in industrialized economies. In the interest of strengthening the statistical evidence, therefore, it would be advisable to extend the time series back in time. However, this can be hardly accomplished in the case under study, since the transition process to capitalism started only in 1989. As such, we can argue that this is a reasonable time period over which to check the achievement of fiscal sustainability in the market economy.

Although the empirical evidence suggests that Poland has progressed towards debt stabilization, the government has the tendency to spend more than its revenues. Using the terminology of Quintos (1995), the budget deficit is sustainable in the long-run, although not in the strong sense: Rather, our results point to a *weak* form of deficit sustainability, which constitutes a warning signal to Polish authorities since the ability of the government to market its debt may not be granted³².

These main findings are obtained, in a first step, within the classical inferential setting. They are supported and more precisely assessed, in a second step, by the adoption of a Bayesian approach to cointegration, through the posterior analysis of the cointegrating vector between revenues and (inclusive of interest) expenditures.

³²This is especially true for developing and emerging countries: Reinhart *et al.* (2003), for instance, argue that emerging economies have a limited capacity for carrying public debt.

References

- [1] Apergis N., Katrakilidis K.P., Tabakis N.M. (2000). Current account deficit sustainability: The case of Greece. *Applied Economics Letters*, **7**: 599-603.
- [2] Arghyrou M.G., Luintel K.B. (2007). Government solvency: Revisiting some EMU countries. *Journal of Macroeconomics*, **29**: 387-410.
- [3] Bauwens L., Giot P. (1998). A Gibbs sampler approach to cointegration. *Computational Statistics*, **13**: 339-368.
- [4] Bauwens L., Lubrano M. (1996). Identification restrictions and posterior densities in cointegrated Gaussian VAR systems. In: Fomby T.B. and Hill R.C. (Eds.), *Advances in Econometrics, Vol 11B Bayesian Methods Applied to Time Series Data*, pp. 3-28, JAI Press, Greenwich.
- [5] Bauwens L., Lubrano M., Richard J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.
- [6] Bauwens L., Richard J.-F. (1985). A poly-t random variable generator, with application to Monte Carlo integration. *Journal of Econometrics*, **29**: 19-46.
- [7] Box G.E.P., Tiao G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading MA.
- [8] Budina N., van Wijnbergen S. (1997). Fiscal policies in Eastern Europe. *Oxford Review of Economic Policy*, **3**: 47-64.
- [9] Buiter W.H. (1996). Aspects of fiscal performance in some transition economies under fund-supported programs. Centre for Economic Policy Research Discussion Paper No. 1535.
- [10] Buiter W.H. (2005). Joys and pains of public debt. In: de Gijsel P. and Schenk H. (Eds.), *Multidisciplinary Economics: The Birth of a New Economics Faculty in the Netherlands*, pp.209-224, Springer.
- [11] Busetti F., Forni L., Harvey A.C., Venditti F. (2007). Inflation convergence and divergence within the European Monetary Union. ECB Working Paper No. 574.
- [12] Caporale G. (1995). Bubble finance and debt sustainability: A test of the government's intertemporal budget constraint. *Applied Economics*, **27**: 1135-1143.
- [13] Chang Y., Park J.Y. (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, **24**: 379-400.
- [14] Chang Y., Park J.Y., Song K. (2006). Bootstrapping cointegrating regressions. *Journal of Econometrics*, **133**: 703-739.
- [15] Corsetti G., Roubini N. (1991). Deficits, public debt and government solvency: Evidence from OECD countries. *Journal of Japanese and International Economics*, **5**: 354-380.
- [16] Davidson R., Flachaire E. (2001). The Wild bootstrap, tamed at last. Queen's Economics Department Working Paper.
- [17] Davis P.J., Rabinowitz P. (1975). *Methods of Numerical Integration*. Academic Press, New York, NY.
- [18] Demetrescu M., Hassler U. (2007). Effect of neglected deterministic seasonality on unit root tests. *Statistical Papers*, **48**: 385-402.
- [19] Dickey D.A., Bell W.R., Miller R.B. (1986). Unit roots in time series models: Tests and implications. *The American Statistician*, **40**: 12-26.
- [20] Drèze J.H. (1977). Bayesian regression analysis using poly-t densities. *Journal of Econometrics*, **6**: 329-354.
- [21] Engle R.F., Granger C.W.J. (1987). Cointegration and error correction: Representation, estimation, and testing. *Econometrica*, **55**: 251-276.
- [22] Fonseca Marinheiro C. (2006). The sustainability of Portuguese fiscal policy from a historical perspective. *Empirica*, **33**: 155-179.
- [23] Geweke J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, **75**: 121-146.
- [24] Green C.J., Holmes M.J., Kowalski T. (2001). Poland: A successful transition to budget sustainability? *Emerging Markets Review*, **2**: 161-183.
- [25] Greiner A., Koeller U., Semmler W. (2006). Testing the sustainability of German fiscal policy: evidence for the period 1960-2003. *Empirica*, **33**: 127-140.

- [26] Haber G., Neck R. (2006). Sustainability of Austrian public debt: a political economy perspective. *Empirica*, **33**: 141-154.
- [27] Hakkio C.S., Rush M. (1991). Is the budget deficit 'too large'? *Economic Inquiry*, **24**: 429-445.
- [28] Hamilton J.D., Flavin M. (1986). On the limitations of government borrowing: A framework for empirical testing. *American Economic Review*, **76**: 808-819.
- [29] Haug A.A. (1991). Cointegration and government borrowing constraints: Evidence for the US. *Journal of Business and Economic Statistics*, **9**: 97-101.
- [30] Hylleberg S., Engle R.F., Granger C.W.J., Yoo B.S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, **44**: 215-238.
- [31] Johansen S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Processes*. 2nd edition, Oxford University Press, Oxford.
- [32] Kalyoncu H. (2005). Fiscal policy sustainability: Test of intertemporal borrowing constraints. *Applied Economics Letters*, **12**: 957-962.
- [33] Kleibergen F., van Dijk H.K. (1994). On the shape of the likelihood/posterior in cointegration models. *Econometric Theory*, **10**: 514-551.
- [34] Kleibergen F., Paap R. (2002). Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics*, **111**: 223-249.
- [35] Koop G., Strachan R.W., van Dijk H.K., Villani M. (2006). Bayesian approaches to cointegration. In: Mills T.C. and Patterson K. (Eds.), *Palgrave Handbook on Econometrics, Vol 1 Econometric Theory*, pp. 871-898, Palgrave MacMillan, Basingstoke.
- [36] Ljungqvist L., Sargent T.J. (2004). *Recursive Macroeconomic Theory*. 2nd edition, The MIT Press, Cambridge MA, London UK.
- [37] MacKinnon J.G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, **11**: 601-618.
- [38] MacKinnon J.G., Haug A., Michelis L. (1999). Numerical distribution functions of the likelihood ratio tests for cointegration. *Journal of Applied Econometrics*, **14**: 563-577.
- [39] Martin G.M. (2000). US deficit sustainability: A new approach on multiple endogenous breaks. *Journal of Applied Econometrics*, **15**: 83-105.
- [40] Ng S., Perron P. (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, **90**: 268-281.
- [41] Palm F.C., Smeekes S., Urbain J.P. (2006). Bootstrap unit root tests: Comparison and extensions. METEOR Research Memorandum 06/015.
- [42] Payne J. (1997). International evidence on the sustainability of budget deficits. *Applied Economics Letters*, **12**: 775-779.
- [43] Psaradakis Z. (2001). Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors. *Journal of Time Series Analysis*, **22**: 577-594.
- [44] Quintos C.E. (1995). Sustainability of the deficit process with structural shifts. *Journal of Business and Economic Statistics*, **13**: 409-417.
- [45] Reinhart C.M., Rogoff K.S., Savastano M.A. (2003). Debt intolerance, NBER Working Paper No. 9908, August.
- [46] Said S.E., Dickey D.A. (1984). Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika*, **71**, 599-607.
- [47] Strachan R.W. (2003). Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics*, **21**: 185-195.
- [48] Strachan R.W., van Dijk H.K. (2003). Bayesian model selection with an uninformative prior. *Oxford Bulletin of Economics and Statistics*, **65**: 863-876.
- [49] Strachan R.W., Inder B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, **123**: 307-325.

- [50] Tompa H. (1973). The iterative Simpson method of numerical integration. CORE Discussion Paper 7336, Université Catholique de Louvain, Belgium.
- [51] Trehan B., Walsh C.E. (1988). Common trends, intertemporal budget balance, and revenue smoothing. *Journal of Economic Dynamics and Control*, **12**: 425-444.
- [52] Trehan B., Walsh C.E. (1991). Testing intertemporal budget constraints: Theory and applications to US federal budget and current account deficits. *Journal of Money, Credit, and Banking*, **23**: 206-223.
- [53] Vanhorebeek F., Van Rompuy P. (1995). Solvency and sustainability of fiscal policies in the EU. *De Economist*, **143**: 457-473.
- [54] Villani M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory*, **21**: 326-357.
- [55] Warne A. (2006). Bayesian inference in cointegrated VAR models with applications to the demand for Euro area M3. ECB Working Paper No. 692.
- [56] Westerlund J., Prohl S. (2007). Testing for fiscal sustainability within the European Union: New evidence based on panel data. Mimeo.
- [57] Wilcox D.W. (1989). The sustainability of government deficits: Implications of the present-value borrowing constraint. *Journal of Money, Credit, and Banking*, **21**: 291-306.

A APPENDIX

A.1 Proof of Result 3.1

The linear model in (6) may be written as

$$\begin{aligned}
\underset{(n \times 1)}{\Delta \mathbf{y}_1} &= \boldsymbol{\Theta} \mathbf{D}_0 + \boldsymbol{\Pi}_1 \Delta \mathbf{y}_0 + \dots + \boldsymbol{\Pi}_{p-1} \Delta \mathbf{y}_{2-p} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_0 + \boldsymbol{\varepsilon}_1 \\
\underset{(n \times 1)}{\Delta \mathbf{y}_2} &= \boldsymbol{\Theta} \mathbf{D}_1 + \boldsymbol{\Pi}_1 \Delta \mathbf{y}_1 + \dots + \boldsymbol{\Pi}_{p-1} \Delta \mathbf{y}_{2-p+1} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_1 + \boldsymbol{\varepsilon}_2 \\
&\vdots \\
\underset{(n \times 1)}{\Delta \mathbf{y}_{T-1}} &= \boldsymbol{\Theta} \mathbf{D}_{T-2} + \boldsymbol{\Pi}_1 \Delta \mathbf{y}_{T-2} + \dots + \boldsymbol{\Pi}_{p-1} \Delta \mathbf{y}_{T-p} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{T-2} + \boldsymbol{\varepsilon}_{T-1} \\
\underset{(n \times 1)}{\Delta \mathbf{y}_T} &= \boldsymbol{\Theta} \mathbf{D}_{T-1} + \boldsymbol{\Pi}_1 \Delta \mathbf{y}_{T-1} + \dots + \boldsymbol{\Pi}_{p-1} \Delta \mathbf{y}_{T-p+1} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{T-1} + \boldsymbol{\varepsilon}_T.
\end{aligned} \tag{34}$$

Note that each equation corresponds to a time observation. Each data point is a $(n \times 1)$ vector, since we are dealing with a multivariate system.

We may further expand (34) as

$$\begin{aligned}
\Delta y_{1,1} &= c_{1,1} + \boldsymbol{\Pi}_1^{(1,1)} \Delta y_{0,1} + \dots + \boldsymbol{\Pi}_1^{(1,n)} \Delta y_{0,n} + \dots + \boldsymbol{\Pi}_{p-1}^{(1,1)} \Delta y_{2-p,1} + \dots + \boldsymbol{\Pi}_{p-1}^{(1,n)} \Delta y_{2-p,n} \\
&\quad + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{0,1} + \varepsilon_{1,1} \\
&\dots \\
\Delta y_{1,n} &= c_{1,n} + \boldsymbol{\Pi}_1^{(n,1)} \Delta y_{0,1} + \dots + \boldsymbol{\Pi}_1^{(n,n)} \Delta y_{0,n} + \dots + \boldsymbol{\Pi}_{p-1}^{(n,1)} \Delta y_{2-p,1} + \dots + \boldsymbol{\Pi}_{p-1}^{(n,n)} \Delta y_{2-p,n} \\
&\quad + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{0,n} + \varepsilon_{1,n} \\
&\vdots \\
\Delta y_{T,1} &= c_{T,1} + \boldsymbol{\Pi}_1^{(1,1)} \Delta y_{T-1,1} + \dots + \boldsymbol{\Pi}_1^{(1,n)} \Delta y_{T-1,n} + \dots + \boldsymbol{\Pi}_{p-1}^{(1,1)} \Delta y_{T-p+1,1} + \dots + \boldsymbol{\Pi}_{p-1}^{(1,n)} \Delta y_{T-p+1,n} \\
&\quad + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{T-1,1} + \varepsilon_{T,1} \\
&\dots \\
\Delta y_{T,n} &= c_{T,n} + \boldsymbol{\Pi}_1^{(n,1)} \Delta y_{T-1,1} + \dots + \boldsymbol{\Pi}_1^{(n,n)} \Delta y_{T-1,n} + \dots + \boldsymbol{\Pi}_{p-1}^{(n,1)} \Delta y_{T-p+1,1} + \dots + \boldsymbol{\Pi}_{p-1}^{(n,n)} \Delta y_{T-p+1,n} \\
&\quad + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{T-1,n} + \varepsilon_{T,n},
\end{aligned} \tag{35}$$

where each row now corresponds to a (1×1) element of the T vectors in (34). With $\boldsymbol{\Pi}_k^{(i,j)}$ we indicate the (i,j) -th entry of the k -th $\boldsymbol{\Pi}$ matrix. The deterministic components in (35) are equal to

$$\begin{aligned}
c_{1,1} &= \Theta_{1,1} D_{0,1} + \dots + \Theta_{1,m} D_{0,m} \\
&\dots \\
c_{1,n} &= \Theta_{n,1} D_{0,1} + \dots + \Theta_{n,m} D_{0,m} \\
&\vdots \\
c_{T,1} &= \Theta_{1,1} D_{T-1,1} + \dots + \Theta_{1,m} D_{T-1,m} \\
&\dots \\
c_{T,n} &= \Theta_{n,1} D_{T-1,1} + \dots + \Theta_{n,m} D_{T-1,m}.
\end{aligned}$$

A.2 Proof of Result 3.2

Defining $\mathbf{E} = (\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B})'(\mathbf{Y} - \tilde{\mathbf{W}}\mathbf{B})$, we can express $\mathbf{E}'\mathbf{E}$ in (15) as a quadratic form in \mathbf{B}

$$\begin{aligned}\mathbf{E}'\mathbf{E} &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{B}'\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}\mathbf{B} \\ &= \tilde{\mathbf{S}} + (\mathbf{B} - \tilde{\mathbf{B}})'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}(\mathbf{B} - \tilde{\mathbf{B}}),\end{aligned}\tag{36}$$

where $\tilde{\mathbf{B}} = (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}$ is the least squares estimator and

$$\tilde{\mathbf{S}} = (\mathbf{Y} - \tilde{\mathbf{W}}\tilde{\mathbf{B}})'(\mathbf{Y} - \tilde{\mathbf{W}}\tilde{\mathbf{B}}).$$

Note that substituting $\tilde{\mathbf{B}} = (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}$ in the expression for $\tilde{\mathbf{S}}$ in (36) we have

$$\begin{aligned}\tilde{\mathbf{S}} &= (\mathbf{Y} - \tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y})'(\mathbf{Y} - \tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}) \\ &= (\mathbf{Y}' - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}')(\mathbf{Y} - \tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}.\end{aligned}\tag{37}$$

Moreover, substituting $\tilde{\mathbf{B}} = (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}$ in the expression $(\mathbf{B} - \tilde{\mathbf{B}})'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}(\mathbf{B} - \tilde{\mathbf{B}})$ in (36) and developing, we get

$$\begin{aligned}&(\mathbf{B} - (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y})'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}(\mathbf{B} - (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}) \\ &= (\mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}} - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{W}})(\mathbf{B} - (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}) \\ &= (\mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}} - \mathbf{Y}'\tilde{\mathbf{W}})(\mathbf{B} - (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y}) \\ &= \mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{Y}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{Y}} \\ &= \mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{Y}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{B}'\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{Y}}.\end{aligned}\tag{38}$$

Finally, summing (37) and (38), we get

$$\begin{aligned}&\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{Y}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{B}'\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{Y}} \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}\mathbf{B} - \mathbf{B}'\tilde{\mathbf{W}}'\mathbf{Y} + \mathbf{B}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}\mathbf{B},\end{aligned}$$

and thus we have proved that $\mathbf{E}'\mathbf{E} = \tilde{\mathbf{S}} + (\mathbf{B} - \tilde{\mathbf{B}})'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}(\mathbf{B} - \tilde{\mathbf{B}})$, as in (36).

A.3 Proof of Result 3.3

We already proved that

$$\begin{aligned}\tilde{\mathbf{S}} &= (\mathbf{Y} - \tilde{\mathbf{W}}\tilde{\mathbf{B}})'(\mathbf{Y} - \tilde{\mathbf{W}}\tilde{\mathbf{B}}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}\tilde{\mathbf{B}} - \tilde{\mathbf{B}}'\tilde{\mathbf{W}}'\mathbf{Y} + \tilde{\mathbf{B}}'\tilde{\mathbf{W}}'\tilde{\mathbf{W}}\tilde{\mathbf{B}}.\end{aligned}$$

Substituting in $\tilde{\mathbf{S}}$ here above the expression $\tilde{\mathbf{B}} = (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{Y}}$, we get

$$\tilde{\mathbf{S}} = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{Y}}$$

and, taking the determinant on both sides,

$$|\tilde{\mathbf{S}}| = |\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{Y}}|.$$

In general, it holds that the determinant of the partitioned matrix $\begin{bmatrix} \tilde{\mathbf{W}}' \tilde{\mathbf{W}} & \tilde{\mathbf{W}}' \mathbf{Y} \\ \mathbf{Y}' \tilde{\mathbf{W}} & \mathbf{Y}' \mathbf{Y} \end{bmatrix}$ is

$$\begin{aligned} & \left| \begin{bmatrix} \tilde{\mathbf{W}}' \tilde{\mathbf{W}} & \tilde{\mathbf{W}}' \mathbf{Y} \\ \mathbf{Y}' \tilde{\mathbf{W}} & \mathbf{Y}' \mathbf{Y} \end{bmatrix} \right| = |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}| |\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}| \\ & \Rightarrow |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} \left| \begin{bmatrix} \tilde{\mathbf{W}}' \tilde{\mathbf{W}} & \tilde{\mathbf{W}}' \mathbf{Y} \\ \mathbf{Y}' \tilde{\mathbf{W}} & \mathbf{Y}' \mathbf{Y} \end{bmatrix} \right| = |\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}| \\ & \Rightarrow |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\mathbf{Y}' \mathbf{Y}| |\tilde{\mathbf{W}}' \tilde{\mathbf{W}} - \tilde{\mathbf{W}}' \mathbf{Y} (\mathbf{Y}' \mathbf{Y})^{-1} \mathbf{Y}' \tilde{\mathbf{W}}| = |\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}| \\ & \Rightarrow |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\mathbf{Y}' \mathbf{Y}| |\tilde{\mathbf{W}}' (\mathbf{I}_T - \mathbf{Y} (\mathbf{Y}' \mathbf{Y})^{-1} \mathbf{Y}') \tilde{\mathbf{W}}| = |\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}| \\ & \Rightarrow |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\mathbf{Y}' \mathbf{Y}| |\tilde{\mathbf{W}}' \mathbf{M}_Y \tilde{\mathbf{W}}| = |\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}|, \end{aligned}$$

where $\mathbf{M}_Y = \mathbf{I}_T - \mathbf{Y} (\mathbf{Y}' \mathbf{Y})^{-1} \mathbf{Y}'$ by definition of the annihilator.

Therefore we conclude that

$$|\tilde{\mathbf{S}}| = |\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \mathbf{Y}| = |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\tilde{\mathbf{W}}' \mathbf{M}_Y \tilde{\mathbf{W}}| |\mathbf{Y}' \mathbf{Y}|. \quad (39)$$

We now need to derive the expressions for $|\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1}$ and $|\tilde{\mathbf{W}}' \mathbf{M}_Y \tilde{\mathbf{W}}|$. Therefore, $|\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|$ is

$$\begin{aligned} |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}| &= \left| \begin{bmatrix} \mathbf{X}' \mathbf{X} & \mathbf{X}' \mathbf{Z} \beta \\ (\mathbf{Z} \beta)' \mathbf{X} & (\mathbf{Z} \beta)' \mathbf{Z} \beta \end{bmatrix} \right| = \left| \begin{bmatrix} \mathbf{X}' \mathbf{X} & \mathbf{X}' \tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}' \mathbf{X} & \tilde{\mathbf{Z}}' \tilde{\mathbf{Z}} \end{bmatrix} \right| \\ &= |\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{Z}}' \tilde{\mathbf{Z}} - \tilde{\mathbf{Z}}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \tilde{\mathbf{Z}}| = |\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{Z}}' (\mathbf{I}_T - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}') \tilde{\mathbf{Z}}| = |\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}|, \end{aligned} \quad (40)$$

being $\mathbf{M}_X = (\mathbf{I}_T - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}')$, by definition. Moreover, coming to $|\tilde{\mathbf{W}}' \mathbf{M}_Y \tilde{\mathbf{W}}|$, we have

$$\begin{aligned} |\tilde{\mathbf{W}}' \mathbf{M}_Y \tilde{\mathbf{W}}| &= \left| \begin{bmatrix} \mathbf{X}' \mathbf{M}_Y \mathbf{X} & \mathbf{X}' \mathbf{M}_Y \mathbf{Z} \beta \\ (\mathbf{Z} \beta)' \mathbf{M}_Y \mathbf{X} & (\mathbf{Z} \beta)' \mathbf{M}_Y \mathbf{Z} \beta \end{bmatrix} \right| \\ &= |\mathbf{X}' \mathbf{M}_Y \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_Y \tilde{\mathbf{Z}} - \tilde{\mathbf{Z}}' \mathbf{M}_Y \mathbf{X} (\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}' \mathbf{M}_Y \tilde{\mathbf{Z}}|. \end{aligned} \quad (41)$$

In the marginal posterior density for β in (19), that we list here below for the sake of the reading,

$$\phi(\beta \mid \text{Data}) \propto \phi(\beta) \times |\tilde{\mathbf{S}}|^{-\frac{(T-k-r)}{2}} |\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-\frac{n}{2}}$$

we can plug in equations (39) and (40). We can obtain

$$\phi(\beta \mid \text{Data}) \propto \phi(\beta) \times \left(|\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\tilde{\mathbf{W}}' \mathbf{M}_Y \tilde{\mathbf{W}}| |\mathbf{Y}' \mathbf{Y}| \right)^{-\frac{(T-k-r)}{2}} \left(|\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}| \right)^{-\frac{n}{2}} \quad (42)$$

Finally, we can substitute for (41) in (42), yielding

$$\begin{aligned} \phi(\beta \mid \text{Data}) &\propto \phi(\beta) \times \left(|\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\mathbf{X}' \mathbf{M}_Y \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_Y \tilde{\mathbf{Z}} - \tilde{\mathbf{Z}}' \mathbf{M}_Y \mathbf{X} (\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}' \mathbf{M}_Y \tilde{\mathbf{Z}}| |\mathbf{Y}' \mathbf{Y}| \right)^{-\frac{(T-k-r)}{2}} \\ &\quad \times \left(|\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}| \right)^{-\frac{n}{2}} \\ &= \phi(\beta) \times \left(|\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\mathbf{X}' \mathbf{M}_Y \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_Y (\mathbf{I}_T - \mathbf{X} (\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}') \mathbf{M}_Y \tilde{\mathbf{Z}}| |\mathbf{Y}' \mathbf{Y}| \right)^{-\frac{(T-k-r)}{2}} \\ &\quad \times \left(|\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}| \right)^{-\frac{n}{2}}. \end{aligned}$$

We set $\tilde{\mathbf{W}}_1 = \tilde{\mathbf{Z}}' \mathbf{M}_Y (\mathbf{I}_T - \mathbf{X} (\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}') \mathbf{M}_Y \tilde{\mathbf{Z}}$ and $\tilde{\mathbf{W}}_2 = \tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}$. We get

$$\begin{aligned} \phi(\beta \mid \text{Data}) &\propto \phi(\beta) \times \left(|\tilde{\mathbf{W}}' \tilde{\mathbf{W}}|^{-1} |\mathbf{X}' \mathbf{M}_Y \mathbf{X}| |\tilde{\mathbf{W}}_1| |\mathbf{Y}' \mathbf{Y}| \right)^{-\frac{(T-k-r)}{2}} \\ &\quad \times \left(|\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{W}}_2| \right)^{-\frac{n}{2}}. \end{aligned} \quad (43)$$

Reminding from (40) that $|\tilde{\mathbf{W}}' \tilde{\mathbf{W}}| = |\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}|$, equation (43) becomes

$$\begin{aligned}
\phi(\boldsymbol{\beta} \mid \text{Data}) &\propto \phi(\boldsymbol{\beta}) \times \left(|\mathbf{X}' \mathbf{X}|^{-1} |\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}|^{-1} |\mathbf{X}' \mathbf{M}_Y \mathbf{X}| |\tilde{\mathbf{W}}_1| |\mathbf{Y}' \mathbf{Y}| \right)^{-\frac{(T-k-r)}{2}} \\
&\quad \times \left(|\mathbf{X}' \mathbf{X}| |\tilde{\mathbf{W}}_2| \right)^{-\frac{n}{2}} \\
&= \phi(\boldsymbol{\beta}) \times \frac{|\mathbf{X}' \mathbf{X}|^{\frac{(T-k-r-n)}{2}} |\tilde{\mathbf{W}}_2|^{\frac{(T-k-r-n)}{2}}}{|\mathbf{X}' \mathbf{M}_Y \mathbf{X}|^{\frac{(T-k-r)}{2}} |\mathbf{Y}' \mathbf{Y}|^{\frac{(T-k-r)}{2}} |\tilde{\mathbf{W}}_1|^{\frac{(T-k-r)}{2}}} \\
&= \phi(\boldsymbol{\beta}) \times k \times \frac{|\tilde{\mathbf{W}}_2|^{\frac{(T-k-r-n)}{2}}}{|\tilde{\mathbf{W}}_1|^{\frac{(T-k-r)}{2}}}, \tag{44}
\end{aligned}$$

where $k = \frac{|\mathbf{X}' \mathbf{X}|^{\frac{(T-k-r-n)}{2}}}{|\mathbf{X}' \mathbf{M}_Y \mathbf{X}|^{\frac{(T-k-r)}{2}} |\mathbf{Y}' \mathbf{Y}|^{\frac{(T-k-r)}{2}}}$ is a term that does not depend on $\boldsymbol{\beta}$. Substituting in (44) the expressions for $\tilde{\mathbf{W}}_2$ and $\tilde{\mathbf{W}}_1$, we get

$$\begin{aligned}
\phi(\boldsymbol{\beta} \mid \text{Data}) &\propto \phi(\boldsymbol{\beta}) \times k \times \frac{|\tilde{\mathbf{Z}}' \mathbf{M}_X \tilde{\mathbf{Z}}|^{\frac{(T-k-r-n)}{2}}}{|\tilde{\mathbf{Z}}' \mathbf{M}_Y (\mathbf{I}_T - \mathbf{X}(\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}') \mathbf{M}_Y \tilde{\mathbf{Z}}|^{\frac{(T-k-r)}{2}}} \\
&= \phi(\boldsymbol{\beta}) \times k \times \frac{|\boldsymbol{\beta}' \mathbf{Z}' \mathbf{M}_X \mathbf{Z} \boldsymbol{\beta}|^{\frac{(T-k-r-n)}{2}}}{|\boldsymbol{\beta}' \mathbf{Z}' \mathbf{M}_Y (\mathbf{I}_T - \mathbf{X}(\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}') \mathbf{M}_Y \mathbf{Z} \boldsymbol{\beta}|^{\frac{(T-k-r)}{2}}} \\
&= \phi(\boldsymbol{\beta}) \times k \times \frac{|\boldsymbol{\beta}' \mathbf{Z}' \mathbf{M}_X \mathbf{Z} \boldsymbol{\beta}|^{l_2}}{|\boldsymbol{\beta}' \mathbf{Z}' \mathbf{M}_Y (\mathbf{I}_T - \mathbf{X}(\mathbf{X}' \mathbf{M}_Y \mathbf{X})^{-1} \mathbf{X}') \mathbf{M}_Y \mathbf{Z} \boldsymbol{\beta}|^{l_1}},
\end{aligned}$$

$l_2 = \frac{T-k-r-n}{2}$ and $l_1 = \frac{T-k-r}{2}$. As usual, $\mathbf{M}_X = \mathbf{I}_T - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$.

Table 1: Descriptive statistics

Variables	Mean	St. dev.	Max	Min	ρ_1 and (Q_1)	ρ_2 and (Q_2)	ρ_{12} and (Q_{12})	ρ_{24} and (Q_{24})
Revenues (log real terms)	5.3011	0.1363	5.5828	4.9551	0.513 (32.318)	0.381 (50.295)	0.253 (124.340)	0.119 (129.510)
Expenditures (log real terms)	5.4235	0.1350	5.6936	4.9180	0.666 (54.553)	0.606 (100.070)	0.279 (272.620)	0.145 (316.340)

The top part of the table is for log of deflated total revenues, the bottom part for log of deflated total expenditures. ρ_i and (Q_i) represent, respectively, the autocorrelation coefficient for values of the series i periods apart (order i) and the Ljung-Box Q-statistics at lag i . For all the series the sample goes from 01/1997 until 12/2006 (120 observations).

Table 2: VAR estimation results and diagnostic checks

Lagged values	Revenues	Expenditures
	rev_t	exp_t
a	3.0426 (0.6093) [4.9932]	2.7593 (0.5410) [5.1002]
DUM_t	-0.1411 (0.0480) [-2.9398]	-0.2112 (0.0426) [-4.9543]
$\delta_{12,t}$	0.0835 (0.0313) [2.6733]	0.0783 (0.0276) [2.8232]
rev_{t-1}	0.2601 (0.1020) [2.5500]	0.0188 (0.0906) [0.2071]
rev_{t-2}	0.0266 (0.1015) [0.2617]	0.0964 (0.0901) [1.0710]
exp_{t-1}	0.0798 (0.1133) [0.7047]	0.3238 (0.1006) [3.2192]
exp_{t-2}	0.0580 (0.1085) [0.5343]	0.0570 (0.0964) [0.5916]
R-squared	0.3923	0.5890
Adjusted R-squared	0.3595	0.5668
S.E. equation	0.0922	0.0819
F-statistic	11.944	26.5122
Log-likelihood	117.4507	131.4841
Akaike AIC	-1.8720	-2.1099
Schwarz SC	-1.7077	-1.9455
VAR residual Serial Correlation LM Test		
Lags	(Serial Correlation) LM-Stat	p-value
2	6.3266	0.1760
VAR residual Normality Test		
Lags	Jarque-Bera Stat	p-value
2	2.0949	0.7183

The table shows VAR estimates for the model in (25). Standard errors are in (\cdot) and t-statistics are in $[\cdot]$. The table also reports some summary regression statistics. 'S.E. equation' is the Standard error of the regression, while 'Akaike AIC' and 'Schwarz SC' are Akaike Information Criterion and Schwarz Criterion, respectively. VAR residual serial correlation and normality test results are also displayed. For all the series the sample goes from 01/1997 until 12/2006 (120 observations).

Table 3: Inverted roots of AR characteristic polynomial

AR root value	Modulus
0.5621	0.5621
-0.2084	0.2084
0.1151-0.1470i	0.1867
0.1151+0.1470i	0.1867

This table presents the inverted roots of the characteristic AR polynomial. No root lies outside the unit circle. The estimated VAR satisfies the stability condition.

Table 4: ADF (parametric and bootstrap) t-test results

Variables (levels)	ADF statistics	5 percent MacKinnon CV
Revenues (log real terms)	-2.2044 (2) (*) (2)	-1.9435 (5%) -2.5847 (1%)
Expenditures (log real terms)	-1.6149 (2) () (2)	-1.9435 (5%) -2.5847 (1%)
Variables (levels)	ADF statistics	5th percentile of the Sieve bootstrap distribution
Revenues (log real terms)	-2.2044 () (2)	-2.4515
Expenditures (log real terms)	-1.6149 () (2)	-2.3982

This table shows results based on a specification with deterministic seasonal dummies in the test regression. For ADF statistics, model lags selected using the recursive procedure proposed by Ng and Perron (1995) are given in parenthesis. The null hypothesis is non-stationarity. The one-sided critical values (CV) for ADF test are obtained from MacKinnon (1996), consult the webpage <http://econ.queensu.ca/faculty/mackinnon/numdist/>. The reported 5th percentiles are those of the distributions for the bootstrap ADF test statistics obtained applying the *sieve* bootstrap scheme with 5,000 replications. The asterisk * denotes rejection of the null hypothesis at the 5% level. For all the series the sample goes from 01/1997 until 12/2006 (120 observations).

Table 5: Linear cointegration rank: Trace test results

Null hypothesis	Eigenvalue	Trace Statistic	0.05 Critical Values
$r = 0$ (*)	0.2660	36.7626	12.3209
$r = 1$	0.0022	0.2615	4.1299

Restricted cointegrating coefficients (Johansen's ML estimates)			
	Coefficient	Estimate	(Standard error)
$r = 1$	β_{exp}	0.8638	(0.1262)

The top part of the table shows the results of the Johansen cointegration rank test. The null hypothesis for the trace test is that the number of cointegrating vectors is r , with the alternative of $r+1$. Critical values are obtained from MacKinnon *et al.* (1999). The asterisk (*) denotes rejection of the null hypothesis at the 5% level. The trace test indicates 1 cointegrating equation at 5% significance level. The bottom part of the table displays the restricted cointegrating coefficient (Johansen's ML point estimate). For this estimate, the chosen normalization is $\beta = [-1 \ \beta_{exp}]'$.

Table 6: Posterior results for β_{exp}

Estimation method: deterministic integration	Mean	(Standard deviation)
Case 1) Student ($\beta_0 = 1, s_0 = 1, m_0 = 1.0589, \nu_0 = 12$) prior: Integration interval (-1, 3)		
Prior	1.0000	(0.3073)
Posterior	0.9296	(0.2285)
Case 2) Beta($\iota = 20, \kappa = 20$) prior: Integration interval (-1, 3)		
Prior	1.0000	(0.3123)
Posterior	0.9232	(0.2392)
Case 3) Truncated flat prior: Integration interval (-1, 3)		
Posterior	0.8936	(0.4918)
Estimation method: maximum likelihood	Estimate	(Standard error)
Posterior mode (under an untruncated flat prior)	0.8006	(0.2883)

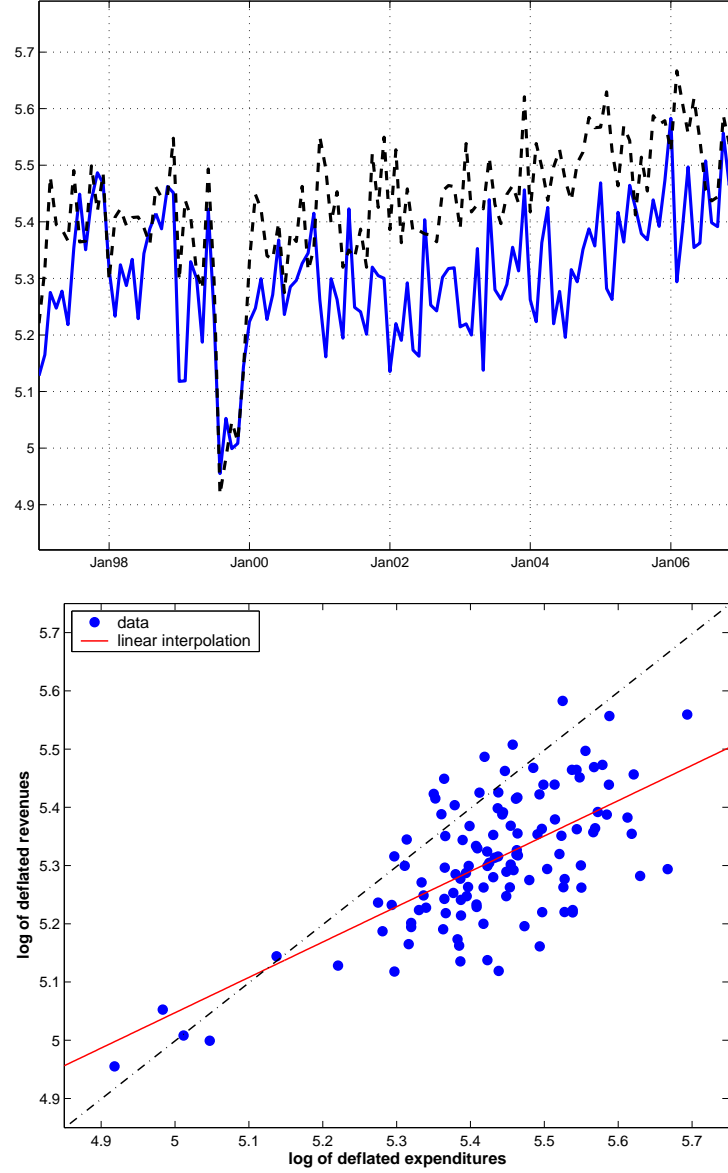
This table presents different estimates of the unrestricted element of the cointegrating vector (β_{exp}). The part part of the table shows Bayesian estimation results, i.e. prior and posterior moments, for the unrestricted element of the cointegrating vector (β_{exp}). 'Prior interval' corresponds to the integration range for numerical integration. Note that the displayed prior moments are exact only for untruncated priors. However, with truncated priors on the (-1, 3) interval, they are close to the exact ones. The bottom part of the table displays the maximum likelihood estimate of the unrestricted element of the cointegrating vector (β_{exp}). The prior is assumed untruncated flat, hence the marginal posterior density is integrable but without finite moments. The maximum likelihood posterior mode and the corresponding asymptotic standard error (Sandwiched standard error) have been calculated using the MATLAB function for optimization Max_lik.m.

Table 7: Posterior results for α vector

	α_{rev}	α_{exp}
Posterior mean	0.3899	-0.0561
(Posterior standard deviation)	(0.3463)	(0.1907)

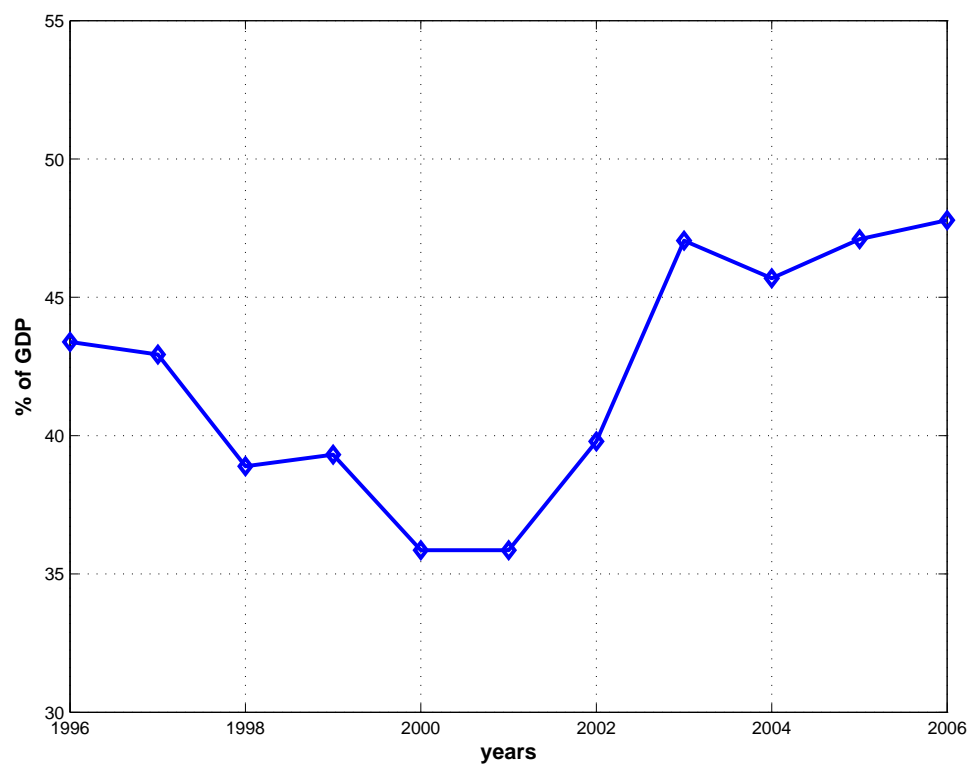
Prior and posterior moments for the adjustment vector.

Figure 1: Log of deflated total revenues and expenditures



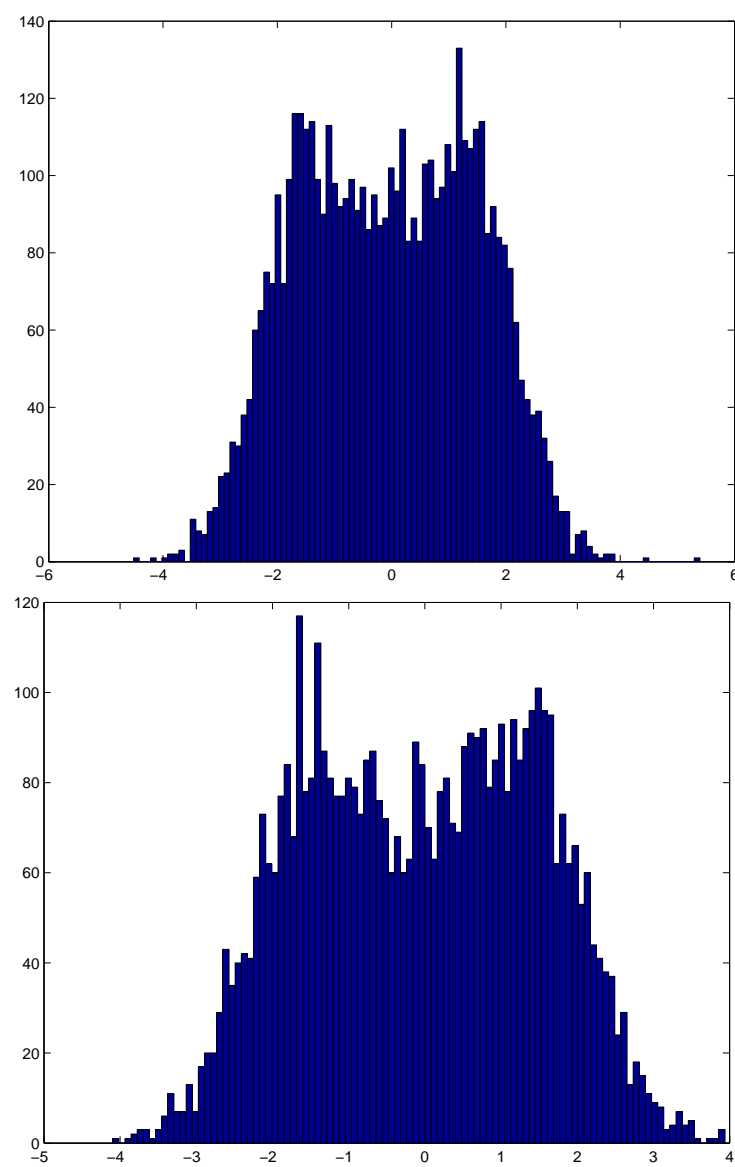
From top to down. First panel: log of deflated total revenues (solid line) and expenditures (dashed line), over the period 1997:01-2006:12. Y-axis is in billions of Polish zlotys. Second panel: cross-plot of log of deflated total revenues and expenditures with 45 degrees line (dash-dotted) and linear interpolation (solid line), whose equation is $\log(rev_t) = 2.010 + 0.607 \times \log(exp_t)$.

Figure 2: Debt-to-GDP ratio



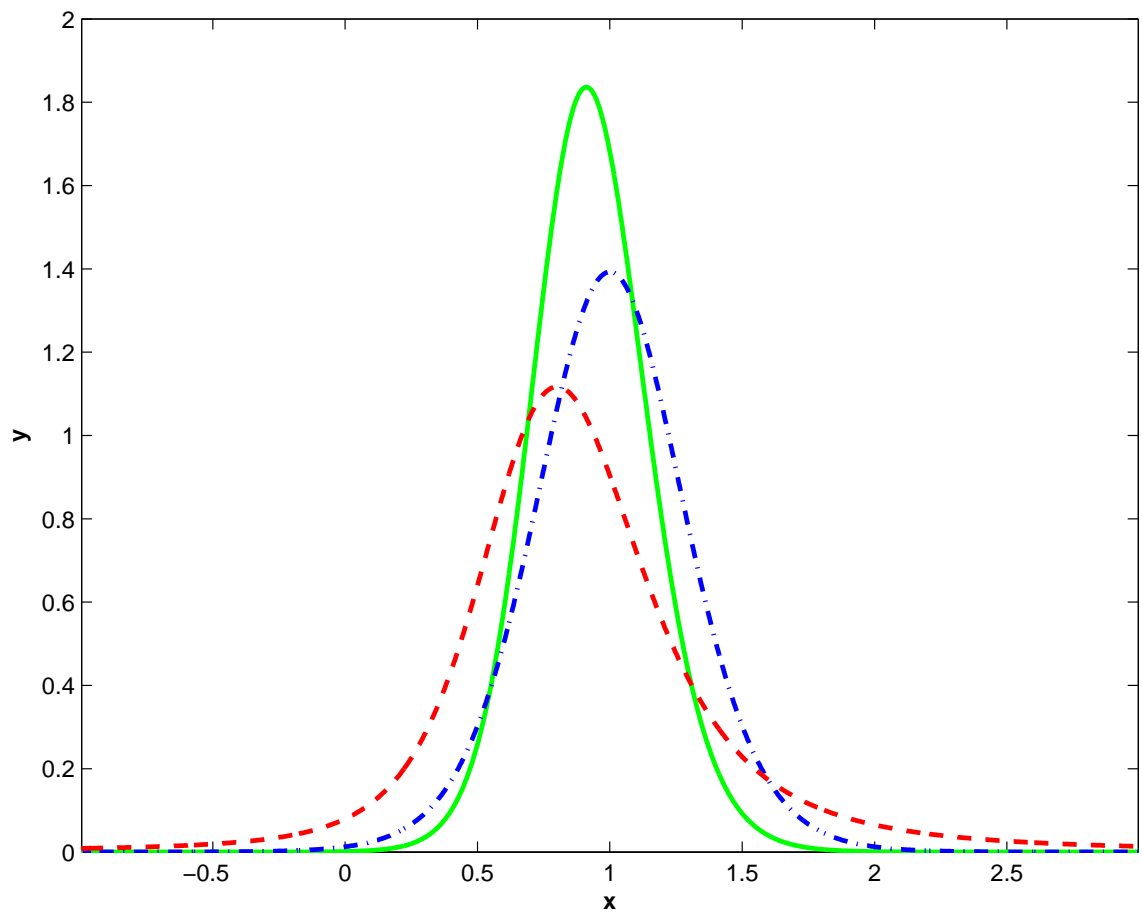
Debt-to-GDP ratio (General Government sector), over the period 1996-2006. The data source is ECB's AMECO.

Figure 3: Sieve bootstrap distributions for ADF t-test statistics



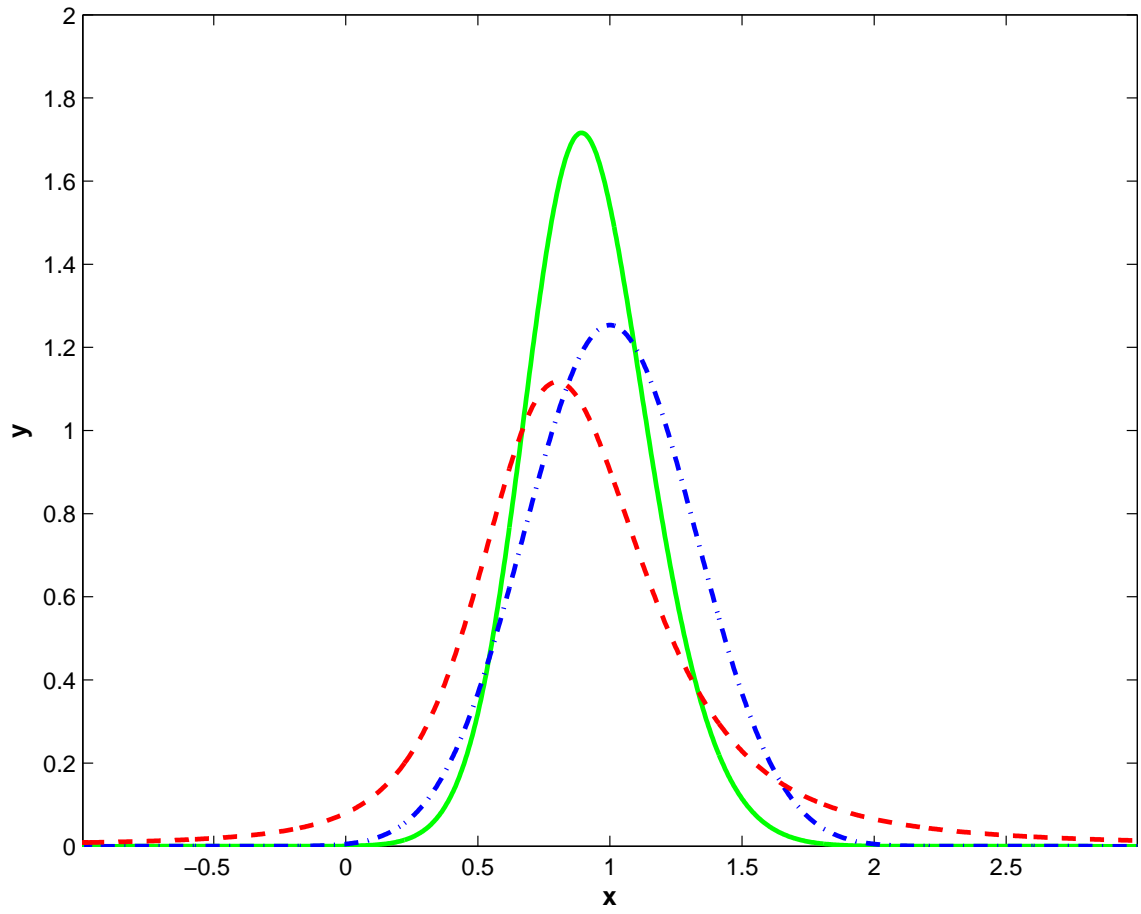
From top to down. First panel: Sieve bootstrap distribution for ADF test statistic for revenues. Second panel: Sieve bootstrap distribution for ADF test statistic for expenditures. The number of replications is 5,000.

Figure 4: Case 1: Posterior density and marginal likelihood for β_{exp} with a truncated Student prior



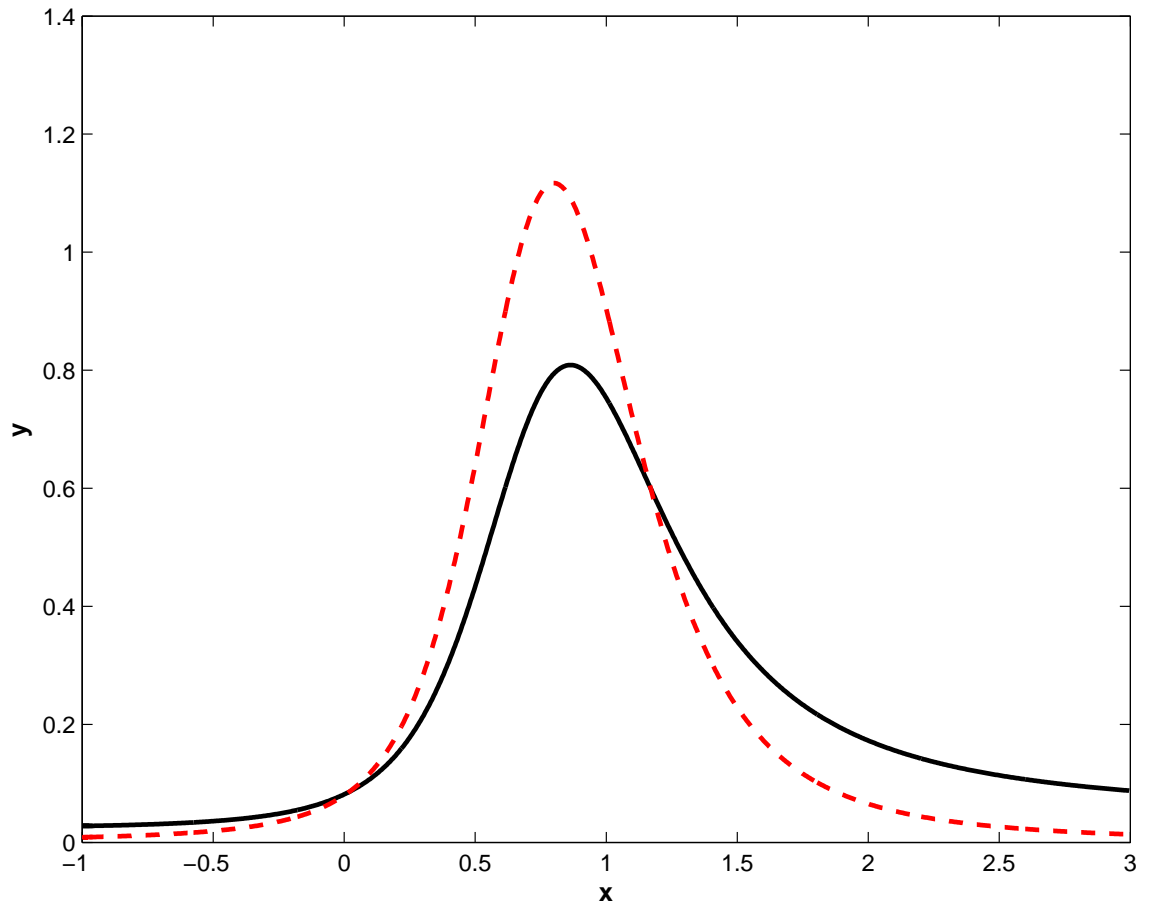
Posterior density (solid line), marginal likelihood function (dashed line) and Student prior (dash-dotted line) for the sampled data. The prior is a Student ($\beta_0 = 1, s_0 = 1, m_0 = 1.0589, \nu_0 = 12$). The integration interval is $(-1, 3)$.

Figure 5: Case 2: Posterior density and marginal likelihood for β_{exp} with a truncated Beta prior



Posterior density (solid line), marginal likelihood function (dashed line) and $\text{Beta}(\iota = 20, \kappa = 20)$ prior (dash-dotted line) for the sampled data. The Beta prior is defined over $(-1, 3)$ and is centred in 1.

Figure 6: Concentrated likelihood function and marginal likelihood function for β_{exp}



Concentrated likelihood function (solid line) and marginal likelihood function (dashed line) for β_{exp} based on the sampled data.